

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 19, 2004, 12:53:28 ; Search time 199.974 Seconds
(without alignments)
10303.209 Million cell updates/sec

Title: US-10-077-130-4_COPY_3662_4146
Perfect score: 485
Sequence: 1 gacgtggtacagcagcgga.....ggagcagctgcacacagga 485

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_29Jan04:*
- 1: Geneseqn1980s:*
 - 2: Geneseqn1990s:*
 - 3: Geneseqn2000s:*
 - 4: Geneseqn2001as:*
 - 5: Geneseqn2001bs:*
 - 6: Geneseqn2002s:*
 - 7: Geneseqn2003as:*
 - 8: Geneseqn2003bs:*
 - 9: Geneseqn2003cs:*
 - 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 485 | 100.0 | 24120 | 7 | ABX11642 |
| 2 | 218.6 | 45.1 | 1251 | 9 | ADC30509 Human nov |
| 3 | 218.6 | 45.1 | 2488 | 4 | AAH18453 Human cDN |
| 4 | 200.4 | 41.3 | 2768 | 7 | ABX71198 Novel hum |
| 5 | 74.4 | 15.3 | 2737 | 4 | AAI59251 Human pol |
| 6 | 74.4 | 15.3 | 3999 | 4 | AAI61037 Human pol |
| 7 | 74.4 | 15.3 | 5382 | 9 | ADD14722 Human src |
| 8 | 73.4 | 15.1 | 1645 | 5 | AAF24162 Human sec |
| 9 | 73.4 | 15.1 | 1948 | 2 | ABV74346 Human IL- |
| 10 | 73.4 | 15.1 | 1949 | 2 | ABV74347 Human IL- |
| 11 | 73.4 | 15.1 | 2170 | 6 | ABQ54970 Human ova |
| 12 | 67.8 | 14.0 | 724 | 4 | AAH08421 Human cDN |
| 13 | 67.8 | 14.0 | 2184 | 4 | AAH18256 Human cDN |
| 14 | 65.8 | 13.6 | 14061 | 6 | ABV99363 Human NOV |
| 15 | 65.8 | 13.6 | 14109 | 6 | ABV99362 Human NOV |
| 16 | 64.2 | 13.2 | 7564 | 7 | ABZ24581 Human cel |
| 17 | 63.6 | 13.1 | 2359 | 7 | ACC46215 Human dit |
| 18 | 61.4 | 12.7 | 2254 | 3 | AAC93403 Human sec |
| 19 | 61.4 | 12.7 | 2354 | 5 | AAI24183 Human sec |
| 20 | 61 | 12.6 | 716 | 6 | ABK35726 cDNA sequ |
| 21 | 57.4 | 11.8 | 572 | 5 | AAI42508 Human cDN |
| 22 | 54.6 | 11.3 | 20565 | 4 | AAK89083 Human dig |
| 23 | 54.6 | 11.3 | 20565 | 4 | AAK89083 Human dig |

| | | | | | |
|----|------|------|-------|---|--------------------|
| 24 | 54.6 | 11.3 | 20565 | 4 | AAI03357 |
| 25 | 54.6 | 11.3 | 20585 | 9 | ADB31749 |
| 26 | 54.2 | 11.2 | 1286 | 6 | ABK35725 |
| 27 | 53.8 | 11.1 | 1005 | 6 | ABN21414 Human ORF |
| 28 | 53.6 | 11.1 | 2884 | 7 | ADA53433 Human cod |
| 29 | 53.2 | 11.0 | 9591 | 4 | AAI28906 Human imm |
| 30 | 53.2 | 11.0 | 9591 | 4 | AAI03355 Human rep |
| 31 | 53.2 | 11.0 | 9591 | 9 | ADB31747 Human nov |
| 32 | 53.2 | 11.0 | 12415 | 4 | AAI28907 Human imm |
| 33 | 53.2 | 11.0 | 12415 | 4 | AAI03356 Human rep |
| 34 | 53.2 | 11.0 | 12415 | 9 | ADB31748 Human nov |
| 35 | 50.6 | 10.4 | 20448 | 4 | ABL19989 Drosophil |
| 36 | 50.6 | 10.4 | 24971 | 4 | ABL19988 Human tit |
| 37 | 49.2 | 10.1 | 81940 | 4 | AAI05390 Human ben |
| 38 | 49.2 | 10.1 | 81940 | 6 | ABK64829 Human RGS |
| 39 | 49.2 | 10.1 | 93801 | 8 | ABX13540 Human RGS |
| 40 | 49 | 10.1 | 3575 | 4 | AAH57402 Human ske |
| 41 | 47.2 | 9.7 | 3597 | 4 | AAI57881 Human pol |
| 42 | 47.2 | 9.7 | 3603 | 4 | AAI59667 Human pol |
| 43 | 46 | 9.5 | 7328 | 4 | AAK89081 Human dig |
| 44 | 44.6 | 9.2 | 352 | 5 | AAF64470 Novel hum |
| 45 | 43.4 | 8.9 | 11916 | 4 | AAH79279 Streptomy |

ALIGNMENTS

RESULT 1
ABX11642
ID ABX11642 standard; cDNA; 24120 BP.

XX AC ABX11642;

XX DT 09-MAY-2003 (first entry)

XX DE Human serine/threonine or protein kinase 12599, cDNA.

XX KW Human; ss; gene; serine/threonine kinase; protein kinase; 12599;
cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
KW haemolytic anaemia; cellular proliferative disorder; cancer;
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.

XX OS Homo sapiens.

XX EH Key Location/Qualifiers

XX FI 5'UTR 1..71

XX FT CDS /*tag= a

XX FT CDS 72..23978

XX FT CDS /*tag= b

XX FT 3'UTR /product= "Kinase 12599"

XX FT 3'UTR /note= "This CDS is specifically claimed in claim 2"

XX FT 3'UTR 23979..24120

XX FT 3'UTR /*tag= c

XX US US2002168742-A1.

XX PD 14-NOV-2002.

XX PF 15-FEB-2002; 2002US-00077130.

XX PR 15-FEB-2001; 2001US-0269201P.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Kapellier-Libermann R, Acton SL;

XX DR WPI; 2003-298729/29.

XX DR P-PSDB; ABG76187.

PT Novel isolated human protein kinase, designated 59079 or 12599
PT polypeptide, useful as diagnostic and therapeutic agents for preventing
PT cardiovascular diseases, proliferative disorders, and protein kinase
PT disorders.
XX
PS Claim 2; Page 58-84; 119pp; English.
XX
XX The invention relates to an isolated human serine/threonine or protein
XX kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule
XX comprising at least 85% identity to the nucleic acids appearing as
XX ABX11641 and ABX11642 or their complement, a naturally occurring variant
XX of the kinases or their fragments. Also included are a non-human host
XX cell containing the nucleic acids, an antibody specific for the proteins,
XX identifying a compound which binds to the kinase (by contacting the
XX kinase or a cell expressing the kinase with a test compound and
XX determining whether the kinase binds to the test compound) and modulating
XX the activity of kinase using the identified compound. The kinases and
XX their encoding nucleic acids are useful as diagnostic and therapeutic
XX agents for preventing a disease or condition associated with an aberrant
XX or unwanted 59079 or 12599 activity in a subject, including
XX cardiovascular diseases such as heart failure, and myocardial infarction;
XX disorders involving blood vessels such as atherosclerosis, and Kaposi's
XX sarcoma; blood platelets disorder such as thrombocytopenia, leukaemia,
XX Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders
XX such as cancer; and protein kinase disorders such as autoimmune
XX disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,
XX rheumatoid arthritis, and multiple sclerosis (many examples of diseases
XX and disorders are included in the specification). The kinases, their
XX encoding nucleic acids and antibodies are useful in screening assays,
XX detection assays (e.g. forensic biology), and predictive medicine (e.g.
XX diagnostic assays, prognostic assays, and monitoring clinical trials and
XX pharmacogenomics). The kinases and their encoding nucleic acids are
XX useful as query sequences to perform a search against public databases to
XX identify other family members or related sequences. The present sequence
XX encodes the kinase 12599
SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;

Query Match 100.0%; Score 485; DB 7; Length 24120;
Best Local Similarity 100.0%; Pred. No. 6.1e-113;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACGTGGTACAGGACGGGAGAGAGCTGAGCTCAGCTCAAAAGTACGATGGAGGTCAA 60
DB 3662 GACGTGGTACAGGACGGGAGAGAGCTGAGCTCAGCTCAAAAGTACGATGGAGGTCAA 3721
QY 61 GGGCTGTCACAGCAAGGCTGGTAGTGCACAGGTGGGCAAGCAGATGCTGGGGAGTACAG 120
DB 3722 GGGCTGTCACAGCAAGGCTGGTAGTGCACAGGTGGGCAAGCAGATGCTGGGGAGTACAG 3781
QY 121 CTGCGAGGCTGGGGCCACAGAGGTCTCTTTCAATGCATCATCAGAGCCCAAGGCGGT 180
DB 3782 CTGCGAGGCTGGGGCCACAGAGGTCTCTTTCAATGCATCATCAGAGCCCAAGGCGGT 3841
QY 181 GTTTCCTCCAGGAGCAGTGGTGCATAATGAGTGGCGACTGAGCAGCGGCGCATGGCCAC 240
DB 3842 GTTTCCTCCAGGAGCAGTGGTGCATAATGAGTGGCGACTGAGCAGCGGCGCATGGCCAC 3901
QY 241 ACTGAGCTGTGAGGTGGGCCCGCCAGCCACAGAGGTGACGCTGGTACAAGGATGGGAAGAA 300
DB 3902 ACTGAGCTGTGAGGTGGGCCCGCCAGCCACAGAGGTGACGCTGGTACAAGGATGGGAAGAA 3961
QY 301 GCTGAGCTCCAGTTCGAAGTCCGATAGAGGCTCGGGCTGCATGGCGAGCTGGTGGT 360
DB 3962 GCTGAGCTCCAGTTCGAAGTCCGATAGAGGCTCGGGCTGCATGGCGAGCTGGTGGT 4021
QY 361 GCAGCAGGACGAGCCAGGACAGATGCTGGGGAGTACACCTGTGAGGCTGGGGGCCAGCGGCT 420
DB 4022 GCAGCAGGACGAGCCAGGACAGATGCTGGGGAGTACACCTGTGAGGCTGGGGGCCAGCGGCT 4081
QY 421 CTCCTTCACCTGGATGTTTCAGAGCCCAAGCGGTGTTTCCAAAGAGCAGCTGGGACA 480
DB 4082 CTCCTTCACCTGGATGTTTTCAGAGCCCAAGCGGTGTTTCCAAAGAGCAGCTGGGACA 4141

QY 481 CAGGA 485
DB 4142 CAGGA 4146
RESULT 2
ADG30509
ID ADC30509 standard; cDNA; 1251 BP.
XX
AC ADC30509;
DT 18-DEC-2003 (first entry)
XX
DE Human novel cDNA sequence, SEQ ID NO:591.
XX
KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerary;
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 17; gene; ss.
XX
OS Homo sapiens.
XX
XX WO2003029271-A2.
XX
PD 10-APR-2003.
XX
PF 24-SEP-2002; 2002WO-US030474.
XX
PR 24-SEP-2001; 2001US-0324631P.
XX
PA (HYSB-) HYSEQ INC.
XX
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX
DR WPI; 2003-371981/35.
DR P-PSDB; ADC31480.
XX
PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
PS Claim 1; SEQ ID NO 591; 1185pp; English.
XX
XX The invention relates to 971 novel human cDNA sequences (ADC29919-
XX ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
XX invention also relates to nucleic acid sequences over 99% identical with
XX the novel human cDNAs. The invention additionally encompasses expression
XX vectors and host cells comprising a nucleic acid of the invention; the
XX recombinant production of a polypeptide of the invention; an antibody
XX against a polypeptide of the invention; a method of detecting
XX polynucleotides or polypeptides of the invention; and methods of
XX identifying a compound which binds to a polypeptide of the invention. The
XX invention further discloses methods of preventing, treating or
XX ameliorating a medical condition; kits comprising polynucleotide probes
XX and/or monoclonal antibodies for carrying out the methods of the
XX invention; methods for the identification of compounds that modulate the
XX expression or activity of the polynucleotide and/or polypeptide; and 767
XX contig sequences corresponding to the cDNA sequences of the invention
XX (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
XX -ADC33394). The nucleic acids and polypeptides of the invention are
XX useful in diagnostics, drug screening, forensics, gene mapping, in the
XX identification of mutations responsible for genetic disorders or other
XX traits, for assessing biodiversity, and in producing many other types of
XX data and products dependent on DNA and amino acid sequences. They are
XX also used for treating diseases such as Parkinson's disease, Alzheimer's

CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primer, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a specifically
CC claimed human cDNA sequence of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 1251 BP; 329 A; 293 C; 363 G; 266 T; 0 U; 0 Other;

Query Match 45.1%; Score 218.6; DB 9; Length 1251;
Best Local Similarity 86.1%; Pred. No. 1.1e-45;
Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 163 CACAGAGCCCAAGGAGTGTGTTGCCAAGGAGCAGTGTGTCATATGAGTGGCGACTGA 222
Db 891 CCCAGAGCCCAAGTGTGTTGCCAAGGAGCAGCACAACAGGAGGTGCGAGGTGA 950

QY 223 GGCAGGGCCAGTGCCCACTGAGCTGTGAGTGGCCAGGCCAGACAGAGGTGAGCTG 282
Db 951 GCGGGGGCCAGTGCCCACTGAGCTGTGAGTGGCCAGGCCAGACAGAGGTGAGCTG 1010

QY 283 GTACAAGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGCTGCGGCTG 342
Db 1011 GTACAAGGATGGGAAGAGCTGAGTTCAGCTCGAAAGTGGCCATAGAGGCTGCGGCTG 1070

QY 343 CATGGGGCAGCTGTGTTGCCAGGAGCGAGCGCCAGCAGATCTGGGAGTACACTGTGA 402
Db 1071 CACAGGAGGTGTGTTGCCAGGAGCGAGCGCCAGCAGATCTGGGAGTACACTGTGA 1130

QY 403 GGCTGGGGCCAGCGGCTCTCTTCCACCTGGATGTTTCAG 443
Db 1131 GGCAGGGGTGTCAGCAGCTCTCTTCCGCTGCGAGTGGCAG 1171

RESULT 3
AAH18453
ID AAH18453 standard; cDNA; 2488 BP.
XX
AC AAH18453;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:18549.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
XX
PR 27-AUG-1999; 99JP-00300253.
XX
PR 11-JAN-2000; 2000JP-00118776.
XX
PR 02-MAY-2000; 2000JP-00183767.
XX
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or

PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 18549; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2488 BP; 561 A; 637 C; 716 G; 574 T; 0 U; 0 Other;

Query Match 45.1%; Score 218.6; DB 4; Length 2488;
Best Local Similarity 86.1%; Pred. No. 1.3e-45;
Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 163 CACAGAGCCCAAGGAGTGTGTTGCCAAGGAGCAGTGTGTCATATGAGTGGCGACTGA 222
Db 891 CCCAGAGCCCAAGTGTGTTGCCAAGGAGCAGCACAACAGGAGGTGCGAGGTGA 950

QY 223 GGCAGGGCCAGTGCCCACTGAGCTGTGAGTGGCCAGGCCAGACAGAGGTGAGCTG 282
Db 951 GCGGGGGCCAGTGCCCACTGAGCTGTGAGTGGCCAGGCCAGACAGAGGTGAGCTG 1010

QY 283 GTACAAGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGCTGCGGCTG 342
Db 1011 GTACAAGGATGGGAAGAGCTGAGTTCAGCTCGAAAGTGGCCATAGAGGCTGCGGCTG 1070

QY 343 CATGGGGCAGCTGTGTTGCCAGGAGCGAGCGCCAGCAGATCTGGGAGTACACTGTGA 402
Db 1071 CACAGGAGGTGTGTTGCCAGGAGCGAGCGCCAGCAGATCTGGGAGTACACTGTGA 1130

QY 403 GGCTGGGGCCAGCGGCTCTCTTCCACCTGGATGTTTCAG 443
Db 1131 GGCAGGGGTGTCAGCAGCTCTCTTCCGCTGCGAGTGGCAG 1171

RESULT 4
ABX71198
ID ABX71198 standard; cDNA; 2768 BP.
XX
AC ABX71198;
XX
DT 05-MAR-2003 (first entry)
XX
DE Novel human cDNA sequence #423.
XX
XX Human; gene; ss; nervous system disorder; peripheral neuropathy;
XX Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
XX neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
XX autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
XX insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
XX ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
XX fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
XX coagulation disorder; cancer; tumour; inflammatory disease; septic shock;

KW Crohn's disease; anaphylaxis; proliferation; chemotactic;
 KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
 KW haemostatic; antiinflammatory; expressed sequence tag; EST.
 XX
 OS Homo sapiens.
 XX
 PN WO200281731-A2.
 PD
 PD 17-OCT-2002.
 XX
 PF 29-JAN-2002; 2002WO-US001222.
 XX
 PR 30-JAN-2001; 2001US-00774528.
 XX
 XX (HYSE-) HYSEQ INC.
 PA (GOOD/) GOODRICH R W.
 XX
 XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
 XX
 XX WPI; 2003-058563/05.
 DR
 XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid
 PT or lymphoid cell disorders, bone disorders, mechanical and traumatic
 PT disorders, coagulation disorders, and inflammatory diseases.
 XX
 XX Claim 1; Page; 612pp; English.
 XX
 CC This invention relates to the cDNA sequences encoding an isolated novel
 CC human polypeptide. The protein encoded by the nucleic acid of the
 CC invention is useful for treating central and peripheral nervous system
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
 CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia)
 CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
 CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
 CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
 CC bacterial, viral or fungal infections; allergic conditions such as
 CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
 CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
 CC disease, anaphylaxis). The protein may be used to inhibit the growth,
 CC infection or function of infectious agents such as bacteria, fungi,
 CC viruses, or to effect bodily characteristics, biorhythms or circadian
 CC cycles of rhythms. The protein may also have
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis
 CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,
 CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
 CC activities. The cDNA sequences of the invention are useful for expressing
 CC recombinant protein for analysis. The present sequence represents a novel
 CC human cDNA sequence of the invention, this sequence is an expressed
 CC sequence tag (EST) and was identified using subtractive hybridisation
 XX
 SQ Sequence 2768 BP; 456 A; 849 C; 995 G; 468 T; 0 U; 0 Other;
 Query Match 41.3%; Score 200.4; DB 7; Length 2768;
 Best Local Similarity 69.8%; Pred. No. 5.4e-41;
 Matches 286; Conservative 0; Mismatches 121; Indels 3; Gaps 1;
 QY 5 TGGTACAGGACGGGAGAACCTCAGCTCCAGCTCAAAAGTACGATCGATGGAGGTCAGGGC 64
 DB 2349 TGGTACAGGATGGATGAGCTGGGACCTCCGGTACGCGCTTCTTCAGGAGGATGTG 2408
 QY 65 TGCACACG--AAGCTGTAGTCAGACAGGTGGGCAAGCAGATGCTGGGGAGTACAGC 121
 DB 2409 GGGACGGCGGACCGGCTGGTGGCAGCCACATCACCAGGCGAGATGAAGGCACCTACTCC 2468
 QY 122 TGGCAGGCTGGGGCCCGCAGAGTCTCTTCACTGACATCAGAGCCCAAGGCGATG 181
 DB 2469 TGGCGGTGGGCGAGGACTCTGTGGACTTCGGGCTCCGGCTCTCTGAGCCCAAGGCGGTG 2528
 QY 182 TTTCACAGGACGAGTGGTGGTCATATAGTGGTGGGACTGAGGCGAGGGGGCCAGTGCACA 241

DB 2529 TTTCACAGGACGAGCGCGCTGCAGGGAGGTGCAGGTGAGTGGGGCCAGCGCCAGC 2588
 QY 242 CTGAGCTGTGAGGTGGCCCGCCAGCCAGAGGTGACGTGTGTAAGGATGGGAAGAG 301
 DB 2589 CTGAGCTGTGAGGTGGCCCGCCAGCCAGATGAGGAGTGCAGTGTGTAAGGATGGGAAGAG 2648
 QY 302 CTGAGCTCCAGTTCCAAAGTGGCATAGAGGCTGGGGGTGTCATCGGCAGCTGGTGGTG 361
 DB 2649 TTGAGCTCCAGTCCGAAAGTGCAGCTGGAGGCTGTGGGCTGTATCGGAGGCTGGTGGTG 2708
 QY 362 CAGCAGGAGGCGCCAGCAGATCTGGGGAGTACCTGTGAGGCTGGGGG 411
 DB 2709 CAGCAGGTCGGCCAGCAGACTCCGATAGTACAGCTGTGAAGCCAGGTG 2758
 RESULT 5
 AAI59251
 ID AAI59251 standard; cDNA; 2737 BP.
 XX
 AC AAI59251;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1454.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 XX 26-JUL-2001.
 XX
 XX 26-DEC-2000; 2000WO-US034263.
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; AAM40095.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 PS Claim 1; SEQ ID NO 1454; 10078pp; English.
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM39642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

| | |
|------------|--|
| CC | lateral sclerosis, and Shy-Drager Syndrome. Other uses include the |
| CC | utilisation of the activities such as; Immune system suppression, |
| CC | Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic |
| CC | and thrombolytic activity, cancer diagnosis and therapy, drug screening, |
| CC | assays for receptor activity, arthritis and inflammation, leukaemias and |
| CC | C.N.S disorders. Note: The sequence data for this patent did not form |
| CC | part of the printed specification |
| XX | |
| SQ | Sequence 2737 BP; 484 A; 840 C; 899 G; 514 T; 0 U; 0 Other; |
| | |
| | Query Match 15.3%; Score 74.4; DB 4; Length 2737; |
| | Best Local Similarity 49.5%; Pred. No. 4.8e-09; |
| | Matches 220; Conservative 0; Mismatches 221; Indels 3; Gaps 1; |
| | |
| QY | 5 TGGTACAGACCGGAGAGCTGAGTGCAGCTCAAAAGTACGCATGGAGGTCARAGGCC 64 |
| DB | 908 TGTGTACAAGACCGGACGAGGAGTGAGAGAGTAGCTTCGTGTGGAGAAATGAGGGG 967 |
| QY | 65 TGCACACGAAGGCTGTGATGTGCAGCAGGTGGGCAAAGCAGATGTGGGAGTACAGCTGC 124 |
| DB | 968 CCCCATGCCCGCTGGTCTGCCGCCACCAGCCCTCACACGGGGCGAGTTTCAGTGC 1027 |
| QY | 125 GAGCTGTGGGGCCAGAGAGTCTCTTTCAACTGCAANTCAGAGCCCCAAGGCAGTGT 184 |
| DB | 1028 GTCGCTGGAGATGAGTGTGCCTACTTTCACGTGCACATCAGACGCTCTCTCGTGGATC 1087 |
| QY | 185 GCCAAGGAGCAGTTGGTGTGCTAATGAGTGTGGGACTGAGCAGGGGCCAGTGCACACTG 244 |
| DB | 1088 GT---GTATCCAGCGCCAGTGTATGTGGACCGCTGGCGCTGTGGTGTCTG 1144 |
| QY | 245 AGCTGTGAGTGGCCCAGGCCACAGAGAGGTGACGTGGTACAAAGNATGGGAAGAAGCTG 304 |
| DB | 1145 ACCTGTGAGCTATGTCGGCCCTGGGCAGAGGTGCGCTGGACCAAGGATGAGAGSAGGTG 1204 |
| QY | 305 AGTCTCAGTTCGAAGTGCATAGAGGCTGCGGCTGCATGCGGACGTGTGTGTGCAG 364 |
| DB | 1205 GTGAGAGCCCCCGCTCTCTTCGAAAGAAAGACACTGTTCGCCGCCCTGGTGTGCTGCC 1264 |
| QY | 365 CAGCAGGCCAGGCAGAGTGTGGGGAGTACACCTGTGAGGCTGGGGGCCAGCGGCTCTCC 424 |
| DB | 1265 GCTGTCCAGCTCGAGGACTCGGGGAGTACTTGTGTAATTGACGATGAGTCGGCTCC 1324 |
| QY | 425 TTCACCTGGATGTTTCAGAGCCC 448 |
| DB | 1325 TTCACTGTCCCGTCACAGAACCC 1348 |
| | |
| RESULT 6 | |
| AAI61037/C | |
| ID | AAI61037 standard; cDNA; 3999 BP. |
| XX | |
| AC | AAI61037; |
| XX | |
| DT | 22-OCT-2001 (first entry) |
| XX | |
| DE | Human polynucleotide SEQ ID NO 5026. |
| XX | |
| KW | Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; |
| KW | peripheral nervous system; neuropathy; central nervous system; CNS; |
| KW | Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; |
| KW | amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; |
| KW | chemokinetic; thrombolytic; drug screening; arthritis; inflammation; |
| KW | leukaemia; ss. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| FN | WO200153312-A1. |
| XX | |
| PD | 26-JUL-2001. |
| XX | |
| FF | 26-DEC-2000; 2000MO-US034263. |
| XX | |
| PR | 23-DEC-1999; 99US-00471275. |

CC The present invention relates to 26 secreted human proteins. The proteins
CC may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, they
CC may be used in gene therapy or in vaccines. Typical of diseases which are
CC potentially treatable are cancers (including leukemia), autoimmune
CC diseases, allergies, inflammation, graft rejection, hyperproliferation,
CC cardiovascular diseases (particularly critical limb ischemia and coronary
CC disease) and any involving abnormal angiogenesis, neurodegeneration
CC and/or infectious diseases
XX
XX

Sequence 1645 BP; 338 A; 474 C; 506 G; 327 T; 0 U; 0 Other;

Query Match 15.1%; Score 73.4; DB 5; Length 1645;
Best Local Similarity 49.4%; Pred. No. 7.7e-09;
Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 5 TGGTACAGAGCGGGAAGAACCTGAGCTCCAGCTCAAAAGTACGCATGGAGGTCAAGGGC 64
Db 905 TGGTACAGAGCGGGAAGAACCTGAGCTCCAGCTCAAAAGTACGCATGGAGGTCAAGGGC 964
QY 55 TGCACAGAGGCTGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGGAGTACAGCTGC 124
Db 965 CCCCATCCGCCCTGTGCTGCTCCGCCCAACCCATCCCTCAGACGGGGGCGAGTTTCAGTGC 1024
QY 125 GAGGCTGGGGCCAGAGAGTCTCTTTCAACTGCACATCAGAGCCCCAAGGCGAGTGT 184
Db 1025 GTCCGTGAGATGAGTGTGCTCTACTTCACTGTCCACATCAGAGCTCTCTCTGTGATC 1084
QY 185 GCCAAGAGACGATGGTGTGATATGAGTGGGACTGAGGAGGGCCAGTGCACACTG 244
Db 1085 GT--GTATCCAGCGGCAAGGTGTATGTGCAGCCGTGCGCTGGAGCGGTGTGCTG 1141
QY 245 AGCTGTAGGTGGCCAGGCCAGACAGAGTGTGCTGTTCAAGGATGGAAGAAGCTG 304
Db 1142 ACCGTGAGCTATGCGCGCTTGGCAGAGTGTGCTGTCGACCAAGGATGGAGAGGTG 1201
QY 305 AGCTCCAGTTGAAAGTCCGATAGAGGTGCGGCTGCATGCGGAGCTGTGTGTGAG 364
Db 1202 GTGGAGAGCCCGCGCTGCTCTGTCAGAAAGACACTGTCCCGCGCTGTGTGCTGCC 1261
QY 365 CAGGAGCGCCAGCAGATGCTGGGAGTACACCTGTGAGGTGGGGCCAGCGCTTCC 424
Db 1262 GCTGTCCAGCTCGAGGATCCCGCGAGTACTTGTGTGAAATGACGATGAGTCGCGCTCC 1321
QY 425 TTCCACCTGGATGTTTCAGAGCC 447
Db 1322 TTCACTGTACCGTTCACAGAGTC 1344

RESULT 9
ID ABV74346
XX ABV74346 standard; cDNA; 1948 BP.
AC ABV74346;
XX
DT 21-FEB-2003 (first entry)
XX
DE Human IL-1RRP3 encoding cDNA SEQ ID NO 1.

XX Human; interleukin 1 receptor related protein 3; IL-1RRP3; inflammation;
KW arthritis; septicemia; autoimmune disease; transplant rejection;
KW infection; stroke; AIDS; osteoporosis; cancer; Alzheimer's disease;
KW anti-inflammatory; antiarthritic; antibacterial; immunosuppressive;
KW cerebroprotective; vasotropic; antiviral; osteopathic; cytostatic;
KW antiarteriosclerotic; neurotropic; neuroprotective; chromosome 2q33; gene;
ss.

OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 413..1699
FT CDS /*tag= a
FT /*product= "IL-1RRP3"

XX WO9918990-A1.
PN 22-APR-1999.
XX
XX 09-OCT-1998; 98WO-US021347.
PF 10-OCT-1997; 97US-0061720P.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
PA
XX Young PR;
XX
XX WPI; 1999-277442/23.
DR P-PSDB; ABB98343.
XX
XX Novel interleukin 1 receptor related protein 3 (IL-1RRP3) or variants
XX for, e.g. treatment of Alzheimer's disease.
XX
XX Claim 2; Page 30-31; 43pp; English.

CC The invention relates to polynucleotides encoding interleukin 1 receptor
CC related protein 3 (IL-1RRP3) or variants. The polynucleotides are used to
CC diagnose a disease or susceptibility to a disease in a subject related to
CC expression or activity of the IL-1RRP3 to screen for compounds which
CC stimulate or inhibit the function of IL-1RRP3 or to treat a subject in
CC need of enhanced or inhibited IL1RRP3 activity. Specifically mentioned in
CC the specification is the use of the IL-3RRP3 polynucleotide or
CC polypeptide for treatment of chronic and acute inflammation, arthritis,
CC septicemia, autoimmune disease, transplant rejection, graft-versus-host
CC disease, infection, stroke, ischaemia, acute respiratory disease
CC syndrome, restenosis, brain injury, AIDS, bone diseases e.g.
CC osteoporosis, cancer, atherosclerosis and Alzheimer's disease. The
CC present sequence is that of a human IL-1RRP3 encoding cDNA of the
CC invention. The IL-1RRP3 gene is located on chromosome 2q33
XX
XX

Sequence 1948 BP; 397 A; 561 C; 600 G; 386 T; 0 U; 4 Other;
Query Match 15.1%; Score 73.4; DB 2; Length 1948;
Best Local Similarity 49.4%; Pred. No. 8e-09;
Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 5 TGGTACAGAGCGGGAAGAGTGCCTCCAGCTCAAAAGTACGCATGGAGGTCAAGGGC 64
Db 1139 TGGTACAGAGCGGGAAGAGTGCCTCCAGCTCAAAAGTACGCATGGAGGTCAAGGGC 1198
QY 65 TGCACACGAGAGGTGTGTAGTGCAGAGGTGGGCAAGCAGATGCTGGGAGTACAGCTGC 124
Db 1199 CCCCATCGCGCGCTGTGTGCTGCCGCCACCCAGCCCTCAGACGGGGCGAGTTTCAGTGC 1258
QY 125 GAGGCTGGGGCCAGAGAGTCTCTTTCAACTGCACATCAGAGAGCCCAAGGAGTGT 184
Db 1259 GTCCGTGGAGATGAGTGTGCTTCACTGTCCATCAGACAGCTCTCTCTGTGTATC 1318
QY 185 GCCAAGGAGCAGATTGTGTGCATAATGAGGTGCGGAGTGCAGGCGGCGCATGTGCCACTG 244
Db 1319 GT---GTATCCAGCGCAGAGTGTATGTGGCAGCGCTGCGCTGGAGCGTGTGTGCTG 1375
QY 245 AGCTGTGAGTGGCCAGAGCCAGAGAGTGTGCTGTTACAGAGTACGAGTGGAGAGAGCTG 304
Db 1376 ACCTGTGAGTATGTCGGGCCCTGGGAGAGTGTGCTGGACCAAGGATGAGAGGAGGTG 1435
QY 305 AGCTCCAGTTGCAAGAGTGGCCATAGAGGCTGCGGGTGTGATGCGGAGCTGGTGGTGCAG 364
Db 1436 GTGGAGAGCCCGCGCTGCTCTTCGAGAGAGAGACACTGTCCGCCCGCTGGTGTGCC 1495
QY 365 CAGCAGGCCAGGAGATGTGTGGGAGTACACCTGTGAGGCTGGGGCCAGCGGCTCTCC 424
Db 1496 GCTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAAATGACGATGATGCGGCTCC 1555
QY 425 TTCCACCTGGATGTTTCAGAGCC 447
Db 1556 TTCACCTGCCGTCACAGAGTC 1578

```

RESULT 10
ABV74347
ID ABV74347 standard; cDNA; 1949 BP.
XX AC ABV74347;
XX DT 21-FEB-2003 (first entry)
XX DE Human IL-1RRP3 encoding cDNA SEQ ID NO 3.
XX KW Human; interleukin 1 receptor related protein 3; IL-1RRP3; inflammation;
XX KW arthritis; septicemia; autoimmune disease; transplant rejection;
XX KW infection; stroke; AIDS; osteoporosis; cancer; Alzheimer's disease;
XX KW anti-inflammatory; antiarthritic; antibacterial; immunosuppressive;
XX KW cerebroprotective; vasotropic; antiviral; osteopathic; cytostatic;
XX KW antiarteriosclerotic; nootropic; neuroprotective; chromosome 2q33; gene;
XX ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 414..1700
XX FT /*tag= a
XX FT /product= "IL-1RRP3"
XX PN WO9918990-A1.
XX PD 22-APR-1999.
XX PF 09-OCT-1998; 98WO-US021347.
XX PR 10-OCT-1997; 97US-0061720P.
XX PA (SMK ) SMITHKLINE BEECHAM CORP.
XX PI Young PR;
XX DR WPI; 1999-277442/23.
XX DR P-PSDB; ABB98344.
XX PT Novel interleukin 1 receptor related protein 3 (IL-1RRP3) or variants
XX PT for, e.g. treatment of Alzheimer's disease.
XX PS Claim 13; Page 31-32; 43pp; English.
XX CC The invention relates to polynucleotides encoding interleukin 1 receptor
XX CC related protein 3 (IL-1RRP3) or variants. The polynucleotides are used to
XX CC diagnose a disease or susceptibility to a disease in a subject related to
XX CC expression or activity of the IL-1RRP3 to screen for compounds which
XX CC stimulate or inhibit the function of IL-1RRP3 or to treat a subject in
XX CC need of enhanced or inhibited IL1RRP3 activity. Specifically mentioned in
XX CC the specification is the use of the IL-3RRP3 polynucleotide or
XX CC polypeptide for treatment of chronic and acute inflammation, arthritis,
XX CC septicemia, autoimmune disease, transplant rejection, graft-versus-host
XX CC disease, infection, stroke, ischaemia, acute respiratory disease
XX CC syndrome, restenosis, brain injury, AIDS, bone diseases e.g.
XX CC osteoporosis, cancer, atherosclerosis and Alzheimer's disease. The
XX CC present sequence is that of a human IL-1RRP3 encoding cDNA of the
XX CC invention. The IL-1RRP3 gene is located on chromosome 2q33
XX SQ Sequence 1949 BP; 398 A; 561 C; 600 G; 386 T; 0 U; 4 Other;
Query Match 15.1%; Score 73.4; DB 2; Length 1949;
Best Local Similarity 49.4%; Pred. No. 8e-09;
Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;
QY 5 TGGTACAAAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAAGGGC 64
DB 1140 TGGTACAAAGGACGGGAGGAGGTGGAGGAGTGGTCTGCTGGAGATGAGGGG 1199
QY 65 TGCACACGAAGGCTGGTAGTGCAGCAGGTGGCAAGACAGATGCTGGGAGTACAGCTCC 124

```

Db 1200 CCCATCGCGCGCTGGTGTGCGCGCCACCCAGCCCTCAGACGGGGGCGAGTTTCAGTGC 1259
 QY 125 GAGGCTGGGGGCCAGAGAGTCTCTTTTCAACTGACATCACAGAGCCCAAGGCAGTGTGTTT 184
 Db 1260 GTGCTGGAGATGAGTGTGCTTACTTCACTGTCAACATCACAGACGTCTCTCTCGTGGATC 1319
 QY 185 GCCAAGGACGATTGGTGCATTAATGAGGTGCGGACTGAGGCGAGGGCCAGTCCACACTG 244
 Db 1320 GT---GTATCCCAGCGGCAAGGTGTATGTGGCAGCGCTGCGCTGTGAGCGTGTGCTGTG 1376
 QY 245 AGCTGTGAGGTGGCCCGCCAGGCCAGACAGAGGTGACGTGTGTAACAGATGGGAAGAGCTG 304
 Db 1377 ACTGTGAGCTATGCGGCGCTTGGCAGAGGTGCGCTGGACCAAGATGGAGAGAGGTG 1436
 QY 305 AGTCCAGTTCGAAAGTGGGCATAGAGGTGCGGGCTGCAATCGCGCAGCTGGTGTGTGAG 364
 Db 1437 GTGGAGAGCCCGCGCTGCTCTCTGCAGAGGAGACACTGTCCGCGCGCTGGTGTGCTGCC 1496
 QY 365 CAGGCAGGCCAGGCAGATGCTGGGAGGTACACTGTGAGGCTGGGGGCCAGGGCTCTCC 424
 Db 1497 GCTGTCCAGCTCGAGGACTCCGCGAGTACTTGTGTGAAATTGACATGATCGGCCCTCC 1556
 QY 425 TTCCACCTGGATGTTTCAGAGCC 447
 Db 1557 TTCACTGTCAACGTCACAGAGTC 1579

RESULT 11
 ABQ54970
 ID ABQ54970 standard; cDNA; 2170 BP.
 XX AC ABQ54970;
 DT 22-AUG-2002 (first entry)
 DE Human ovarian antigen HNOCH54 cDNA, SEQ ID NO:850.
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive; gene; ss.
 OS Homo sapiens.
 XX WO200200677-A1.
 XX 03-JAN-2002.
 XX 07-JUN-2001; 2001WO-US018569.
 XX 07-JUN-2000; 2000US-0209467P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Birse CE, Rosen CA;
 XX WPI: 2002-147878/19.
 XX P-PSDB; ABF41893.
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 PT cancer), immune disorders, cardiovascular disorders and neurological
 PT diseases.
 PS Claim 1; SEQ ID NO 850; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also encompasses polypeptides 90% identical and polynucleotides 95% identical to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human ovarian antigen polynucleotides, antibodies against human ovarian antigens, and the use of ovarian antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various ovary and/or breast-related disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), blood-related disorders (e.g., anaemia), cardiovascular disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents cDNA encoding a human ovarian antigen of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/bct/sequences

```

Query Match      15.1%; Score 73.4; DB 6; Length 2170;
Best Local Similarity 49.4%; Pred. No. 8.2e-09;
Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;
Sequence 2170 BP; 456 A; 624 C; 679 G; 411 T; 0 U; 0 Other;

```

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|------|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|------|------|------|------|
| QY | 5 | TG | G | T | A | C | A | A | G | C | G | G | A | A | G | A | G | C | T | C | A | A | A | G | T | A | C | G | C | A | A | A | G | T | C | A | A | G | G | C | 64 | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 1396 | TG | G | T | A | C | A | A | G | C | G | G | A | A | G | A | G | C | T | C | T | G | T | G | C | T | G | G | A | A | T | A | C | G | A | T | G | A | G | G | 1455 | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| QY | 65 | TG | C | A | C | A | A | G | G | C | T | G | T | A | G | T | G | C | A | G | G | T | G | G | C | A | A | G | C | A | G | A | T | C | T | C | G | G | A | G | 124 | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 1456 | CC | C | C | A | T | C | G | C | C | T | G | T | G | T | G | C | C | G | G | C | A | C | C | C | C | T | C | A | G | C | G | G | G | G | C | A | G | T | 1515 | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| QY | 125 | G | A | G | C | T | G | G | G | G | C | C | A | G | A | G | T | C | T | C | T | T | C | A | A | C | A | T | C | A | C | A | G | C | C | A | A | G | C | A | G | 184 | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 1516 | G | T | G | C | T | G | G | A | T | G | A | G | T | G | T | G | C | T | A | C | T | G | T | C | A | C | C | A | T | C | A | C | A | G | C | T | C | T | C | G | T | 1575 |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| QY | 185 | G | C | A | A | G | A | G | C | A | G | T | T | G | T | G | C | A | T | A | T | A | G | A | G | T | C | G | G | A | C | T | G | A | G | C | A | G | G | C | A | G | 244 |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Db | 1576 | G | T | - | - | - | G | P | A | T | C | C | C | A | G | C | G | C | A | A | G | T | G | T | G | G | C | A | G | C | C | T | G | C | C | T | G | A | G | C | T | 1632 | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| QY | 245 | A | G | C | T | G | C | A | G | T | G | G | C | C | A | G | C | C | C | A | G | A | G | T | | | | | | | | | | | | | | | | | | | |

RESULT 12
AAH08421
ID AAH08421 standard; cDNA; 724 bp.
XX

| | | |
|----|--|--|
| AC | AAH08421; | |
| XX | | |
| XX | 26-JUN-2001 (first entry) | |
| DT | | |
| XX | | |
| DE | Human cDNA clone (5'-primer) SEQ ID NO:5256. | |
| DE | | |
| XX | Human; primer; detection; diagnosis; anti-sense therapy; gene therapy; ss. | |
| KW | | |
| XX | | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| FN | EP1074517-A2. | |
| XX | | |
| PD | 07-FEB-2001. | |
| XX | | |
| PF | 28-JUL-2000; 2000EP-00116126. | |
| XX | | |
| PR | 29-JUL-1999; 99JP-00248036. | |
| PR | 27-AUG-1999; 99JP-00300253. | |
| PR | 11-JAN-2000; 2000JP-00118776. | |
| PR | 02-MAY-2000; 2000JP-00183767. | |
| PR | 09-JUN-2000; 2000JP-00241899. | |
| XX | | |
| XX | | |
| PA | (HELI-) HELIX RES INST. | |

AC

Db 360 CCTGAAGACGCCGCTGTACCTGCCACGTGGGCTCGAGGAGACCCGGGCCCGGGTC 419
Qy 434 GATTTTCAGAGCCCAAGCGGTGTTTGCAAGAGCAGCTGSCA 478
Db 420 CGCGTGACGATCTGTCACGTGGGCATCACCAGAGGGCTGAAGACA 464

Search completed: September 19, 2004, 15:42:36
Job time : 202.974 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:09:24 ; Search time 43.5558 Seconds
(without alignments)
6179.453 Million cell updates/sec

Title: US-10-077-130-4_COPY_3662_4146
Perfect score: 485
Sequence: 1 gacgtgtacaggaagcggga.....ggagcagctggcacacagga 485

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----------------------|-------------------|
| C 1 | 63.4 | 13.1 | 7218 | 1 US-08-232-463-14 | Sequence 14, Appl |
| 2 | 47.2 | 9.7 | 3575 | 4 US-09-023-655-1395 | Sequence 1395, Ap |
| 3 | 41 | 8.5 | 5207 | 4 US-09-858-664A-1 | Sequence 1, Appl |
| 4 | 41 | 8.5 | 5207 | 4 US-10-274-978-1 | Sequence 1, Appl |
| 5 | 41 | 8.5 | 5207 | 4 US-10-274-978-3 | Sequence 3, Appl |
| 6 | 40.4 | 8.3 | 2793 | 1 US-08-209-747-1 | Sequence 1, Appl |
| 7 | 40.4 | 8.3 | 2793 | 1 US-08-458-298-1 | Sequence 1, Appl |
| 8 | 39.8 | 8.2 | 1548 | 2 US-08-762-106-5 | Sequence 5, Appl |
| 9 | 39.8 | 8.2 | 1548 | 3 US-09-320-774-5 | Sequence 5, Appl |
| 10 | 39.8 | 8.2 | 1581 | 2 US-08-762-106-6 | Sequence 6, Appl |
| 11 | 39.8 | 8.2 | 1581 | 3 US-09-320-774-6 | Sequence 6, Appl |
| C 12 | 39.6 | 8.2 | 5020 | 3 US-08-938-291A-3 | Sequence 3, Appl |
| 13 | 39.6 | 8.2 | 5020 | 4 US-09-589-619-3 | Sequence 3, Appl |
| 14 | 39.4 | 8.1 | 1708 | 4 US-09-919-172-30 | Sequence 30, Appl |
| 15 | 39.4 | 8.1 | 1709 | 2 US-09-010-388-2 | Sequence 2, Appl |
| 16 | 39.4 | 8.1 | 1709 | 3 US-09-366-260-2 | Sequence 2, Appl |
| 17 | 39.4 | 8.1 | 1709 | 3 US-09-153-804-11 | Sequence 11, Appl |
| C 18 | 37.8 | 7.8 | 50937 | 3 US-09-428-517-1 | Sequence 1, Appl |
| 19 | 37.4 | 7.7 | 1926 | 4 US-09-249-585A-2 | Sequence 2, Appl |
| 20 | 37.4 | 7.7 | 1926 | 4 US-09-410-399-3 | Sequence 3, Appl |
| 21 | 37.4 | 7.7 | 2580 | 3 US-09-050-863-2 | Sequence 2, Appl |
| 22 | 37.4 | 7.7 | 2580 | 4 US-09-359-081-2 | Sequence 2, Appl |
| C 23 | 37.4 | 7.7 | 5452 | 2 US-09-130-114-1 | Sequence 1, Appl |
| C 24 | 37.4 | 7.7 | 8705 | 4 US-09-647-344A-14 | Sequence 14, Appl |
| 25 | 37.4 | 7.7 | 9600 | 3 US-08-910-647-1 | Sequence 1, Appl |
| 26 | 37.4 | 7.7 | 9600 | 4 US-09-620-925-1 | Sequence 1, Appl |
| 27 | 37.4 | 7.7 | 10596 | 1 US-07-884-811-15 | Sequence 15, Appl |

| | | | | | |
|------|------|-----|-------|-----------------------|-------------------|
| 28 | 37.4 | 7.7 | 10596 | 1 US-07-885-971-15 | Sequence 15, Appl |
| 29 | 37.4 | 7.7 | 10596 | 1 US-08-087-783A-15 | Sequence 15, Appl |
| 30 | 37.4 | 7.7 | 10596 | 1 US-08-194-088B-15 | Sequence 15, Appl |
| 31 | 37.4 | 7.7 | 10596 | 2 US-08-194-087-15 | Sequence 15, Appl |
| 32 | 37.4 | 7.7 | 10596 | 5 PCT-US93-04648-15 | Sequence 15, Appl |
| C 33 | 37.4 | 7.7 | 16080 | 4 US-09-724-566A-48 | Sequence 48, Appl |
| 34 | 37.4 | 7.6 | 1866 | 3 US-09-318-448-33 | Sequence 33, Appl |
| C 35 | 37 | 7.6 | 1866 | 4 US-09-016-434-1323 | Sequence 1323, Ap |
| 36 | 37 | 7.6 | 5690 | 2 US-08-447-464-2 | Sequence 2, Appl |
| 37 | 37 | 7.6 | 5690 | 2 US-08-716-679-2 | Sequence 2, Appl |
| C 38 | 36.8 | 7.6 | 1392 | 4 US-09-252-991A-8917 | Sequence 8917, Ap |
| 39 | 36.8 | 7.6 | 1500 | 4 US-09-252-991A-9248 | Sequence 9248, Ap |
| 40 | 36.8 | 7.6 | 1956 | 4 US-09-252-991A-9165 | Sequence 9165, Ap |
| 41 | 36.6 | 7.5 | 2793 | 2 US-08-795-868-13 | Sequence 13, Appl |
| 42 | 36.6 | 7.5 | 2793 | 4 US-09-303-069-13 | Sequence 13, Appl |
| 43 | 36.6 | 7.5 | 2293 | 4 US-09-134-250-13 | Sequence 13, Appl |
| 44 | 36 | 7.4 | 2225 | 4 US-09-620-312D-1046 | Sequence 1046, Ap |
| 45 | 36 | 7.4 | 5661 | 4 US-08-938-105-2 | Sequence 2, Appl |

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)893-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 13.1%; Score 63.4; DB 1; Length 7218;

FILE REFERENCE: CL000927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Human
US-10-274-978-1

Query Match 8.5%; Score 41; DB 4; Length 5207;
Best Local Similarity 57.4%; Pred. No. 0.13;
Matches 74; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 1 GACGTGGTACAGGACGGGAGAGCTGAGCTCCAGCTCAAAAGTACGCTAGGAGTCAA 60
DB 363 GACCTGGTACAGGACAGCGTCCAGCTGGTGGACAGACCCGGCTTAGCCAGCAGA 422
QY 61 GGGCTGCACACGAAGGCTGGTAGTCAGCAGGTGGGCAAGACAGATGCTGGGGAGTACAG 120
DB 423 AGGCACACATACCTCCCTGGTGTGAGGCATGTGGCCTCGAAGGATGCCGCGTTTACAC 482
QY 121 CTGCGAGGC 129
DB 483 CTGCTGGC 491

RESULT 5
US-10-274-978-3
; Sequence 3, Application US/10274978
; Patent No. 6670164
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; MOLECULE TYPE: LINEAR
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Human
US-10-274-978-3

Query Match 8.5%; Score 41; DB 4; Length 5207;
Best Local Similarity 57.4%; Pred. No. 0.13;
Matches 74; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 1 GACGTGGTACAGGACGGGAGAGCTGAGCTCCAGCTCAAAAGTACGCTAGGAGTCAA 60
DB 363 GACCTGGTACAGGACAGCGTCCAGCTGGTGGACAGACCCGGCTTAGCCAGCAGA 422
QY 61 GGGCTGCACACGAAGGCTGGTAGTCAGCAGGTGGGCAAGACAGATGCTGGGGAGTACAG 120
DB 423 AGGCACACATACCTCCCTGGTGTGAGGCATGTGGCCTCGAAGGATGCCGCGTTTACAC 482
QY 121 CTGCGAGGC 129
DB 483 CTGCTGGC 491

RESULT 6
US-08-209-747-1
; Sequence 1, Application US/08209747
; Patent No. 5733771
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Colgin, Mark
; TITLE OF INVENTION: CDNAS Encoding Minor Ampullate Spider
; TITLE OF INVENTION: Silk Proteins
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/209,747
; FILING DATE: 14-MAR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2793 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Nephila clavipes
; TISSUE TYPE: minor ampullate gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 183..2675
; OTHER INFORMATION: /product= "N. clavipes minor
; OTHER INFORMATION: ampullate silk protein"
US-08-209-747-1

Query Match 8.3%; Score 40.4; DB 1; Length 2793;
Best Local Similarity 45.8%; Pred. No. 0.15;
Matches 140; Conservative 0; Mismatches 166; Indels 0; Gaps 0;
QY 167 GAGCCCAAGCAGCTGTTGCCAAGGAGCAGTGGTGCATATAGAGTGGAGTGGAGGCA 226
DB 520 GTCCGGAGCCGAGAGCTGGTGTCTGCACTCTCTGAGCAGGAGCCGGAGGCGCTGGTG 579
QY 227 GGGGCCACAGTCCACACTGAGCTGTGAGGTGGCCAGCCAGACAGAGTGCAGTGGTAC 286
DB 580 GTTACGCTACAGTGTCTGGTGTCTGCGAGTGGTCTGGGCGAGGTGCAGGCGCCG 639
QY 287 AAGATCGGAGAGTGCAGTCCAGTTCGAAGTGGCATAGAGCTCGGGCTCATG 346
DB 640 GTGGTGTGATATGTTGGACAAGGGCGATATGTTGCGGAGCAGGAGTGTGGCGCTG 699
QY 347 CGGCAGCTGGTGTGAGCAGGAGGAGGAGTGTGGGAGATGCTGGGAGTACACCTGTGAGGT 406
DB 700 CTGCTGCTGTGTCAGGAGGAGGAGTGTGGCGTTACGCTAGAGTGTGCTGGTGGAG 759
QY 407 GGGGGCCAGCGGCTCTCTTCCACCTGGATGTTTCAGAGCCCAAGGGGTGTTGCAAG 466

QY 302 CTGAGCTCCAGTTGCGAAGTCCGATAGAGGCTCGGGCTGCATCGCGCAGCTGGTGGT 361
Db 1225 CGCGGAGGAGCAGAGGCGCTCTGAGAGAGTTGGCGGCCACGCGCGGAGGCG 1284
QY 362 CAGCAGGAGCAGCGCAGATGCTGGGGAGTACCTGTGAGGCTGGGGGCGCAGCGCTC 421
Db 1285 GAGGAGCGCGGAGGAGCGCCCGCGGAGCGCGCTGCTGCGCGATGCGGTGGAGC-- 1342
QY 422 TCCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTTGCAAGGAGCAGCTGGCACAC 481
Db 1343 -CTGTGACGTGGAGCTGAGTGGCCNAGCGGTGGAGGCTGTGCCNAGCGCGCTC 1401
QY 482 AGG 484
Db 1402 AAG 1404

RESULT 9

US-09-320-774-5
; Sequence 5, Application US/09320774
; Patent No. 6265545
; GENERAL INFORMATION:
; APPLICANT: Jarvik, Jonathan W.
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE
; TITLE OF INVENTION: TAGGING
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harris Brotman
; STREET: 202 Coast Blvd., Suite 111
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09320,774
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/762,106
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brotman, Harris F.
; REGISTRATION NUMBER: 35,461
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 654-2428
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1548 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-320-774-5

Query Match 8.2%; Score 39.8; DB 3; Length 1548;
Best Local Similarity 48.2%; Pred. No. 0.19;
Matches 175; Conservative 0; Mismatches 182; Indels 6; Gaps 2;
QY 125 GAGGCTGGGGCCAGAGAGTCTCCTTCACTGTCATCAGAGCCCAAGCGAGTGT 184
Db 1045 GAGGCGGTGCGCGCGCGCCACGTTGTGTCGCGAGCTCAAGCCCGGTGGCGAGC 1104
QY 185 GCCAA---GGAGCAGTTGGTGATATAGGTGCGGACTGAGGCGAGGCGCCAGTGCACAC 241
Db 1105 GCGATGCCGTGAGGCGCGCGCGGAGTGAACGCGCAGCGCGGAGGCGCCCAAC 1164

QY 242 CTGAGCTGTAGTGGCCCGAGCCAGACAGAGGTGACGTGTGTAACAAGGATGGGAAGAAG 301
Db 1165 GCCAAGTGGGAGCGGACAAAGCGGAGCGCGGAGAAAGCGCGCCGAGCGGAGGCG 1224
QY 302 CTGAGCTCCAGTTCGAAAGTGGCATAGAGGCTGCGGGCTGCATGGCGCAGCTGGTGTG 361
Db 1225 GCGCGGAGGAGCAGAGGCGCTGCTGGAGGAGTTGGCGGCCACGCGCGCGCGGAGGCG 1284
QY 362 CAGCAGGAGCGCAGCAGATGCTGGGGAGTACACTGTGAGGCTGGGGGCGCAGCGGCTC 421
Db 1285 GAGGAGCGCGGAGGAGCGCGCGCGGAGCGCGCTGCTGCCGATGGCGTGGAGC-- 1342
QY 422 TCCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTTTGCAGAGGAGCAGCTGGCACAC 481
Db 1343 -CTGTGACGTGGAGCTGAGTGGCCNAGCGGTGGAGGCTGTGCCNAGCGCGCTC 1401
QY 482 AGG 484
Db 1402 AAG 1404

RESULT 10

US-08-762-106-6
; Sequence 6, Application US/08762106
; Patent No. 5948677
; GENERAL INFORMATION:
; APPLICANT: Jarvik, Jonathan W.
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE
; TITLE OF INVENTION: TAGGING
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harris Brotman
; STREET: 202 Coast Blvd., Suite 111
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,106
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brotman, Harris F.
; REGISTRATION NUMBER: 35,461
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 654-2428
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-762-106-6

Query Match 8.2%; Score 39.8; DB 2; Length 1581;
Best Local Similarity 48.2%; Pred. No. 0.19;
Matches 175; Conservative 0; Mismatches 182; Indels 6; Gaps 2;
QY 125 GAGGCTGGGGCCAGAGAGTCTCCTTCACTGTCATCAGAGCCCAAGCGAGTGT 184
Db 1078 GAGGCGGTGCGCGCGCGCCACGTTGTGTCGCGAGCTCAAGCCCGGTGGCGAGC 1137
QY 185 GCCAA---GGAGCAGTTGGTGATATAGGTGCGGACTGAGGCGAGGCGCCAGTGCACAC 241

Db 1138 GCCGATGCCGTGAGCGCGCGCCGCGAGCTGACGGCGAGGCGGAGGCGGCCAAC 1197
 QY 242 CTGAGCTGTGAGTGGCCAGCCAGACAGAGTGTGATCAAGGATGGGAAGAAG 301
 Db 1198 GCCAGTGGGAGCGGACAAAGCGGAGCGCGGAGAGGCGCGCGCGGAGGCG 1257
 QY 302 CTGAGCTCCAGTTCGAAAGTGGCAGATGCGGCTGCTGCGGCTGCTGCGGCTG 361
 Db 1258 GCGGCGGAGGAGCAGAGGCGCTGCTGAGGAGTGGCGGCGACGCGCGCGGAGGCG 1317
 QY 362 CAGCAGGCGAGCGAGGAGTGTGGGAGTACACCTGTGAGGCTGGGGCCAGCGCTC 421
 Db 1318 GAGGAGCGGCGGAGGAGCGCGCGGAGCGCGCGCTGCTGCGGATGGCGTGAGC-- 1375
 QY 422 TCCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTTTCAAGAGGAGCAGCTGGCACAC 481
 Db 1376 -CTGTGACGTGGAGGCTGAGGTGGCCAAAGCGGTGGAGGCTGTGCCAAGCGCGGTC 1434
 QY 482 AGG 484
 Db 1435 AAG 1437

RESULT 11

US-09-320-774-6
 ; Sequence 6, Application US/09320774
 ; Patent No. 6265545

; GENERAL INFORMATION:
 ; APPLICANT: Jarvik, Jonathan W.
 ; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE
 ; TITLE OF INVENTION: TAGGING
 ; NUMBER OF SEQUENCES: 47
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harris Brotman
 ; STREET: 202 Coast Blvd., Suite 111
 ; CITY: La Jolla
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/320,774
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/762,106
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brotman, Harris F.
 ; REGISTRATION NUMBER: 35,461
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 654-2428
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1581 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; US-09-320-774-6

Query Match 8.28; Score 39.8; DB 3; Length 1581;

Best Local Similarity 48.24; Pred. No. 0.19;
 Matches 175; Conservative 0; Mismatches 182; Indels 6; Gaps 2;

QY 125 GAGGCTGGGCGCAGAGAGTCTCTTTCAACTGCACATCACAGAGCCCAAGGAGTGT 184
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1078 GAGGCGGTCCGGCGCGGCCACCTTTGTGTGTCGCGAGCTCAAGCCCCGCGTGGCGAGC 1137
 QY 185 GCCAA---GGAGCAGTTTGTGCTAATGAGGTGCGGACTGAGGCGAGGGGCCAGTCCAC 241
 Db 1138 GCCGATGCCGTGAGCGCGCGCCGCGAGCTGACGGCGCAGGCGGAGGAGGCGGCCAAC 1197
 QY 242 CTGAGCTGTGAGTGGCCAGCCAGACAGAGTGTGATCAAGGATGGGAAGAAG 301
 Db 1198 GCCAAGTGGAGCGGACAAAGCGGAGCGCGGAGAGGCGCGCGCGGAGGCG 1257
 QY 302 CTGAGCTCCAGTTCGAAAGTGGCAGATGAGGCTGCGGCTGCTGCGGCTGCTGCGGCTG 361
 Db 1258 GCGGCGGAGGAGCAGAGGCGCTGCTGAGGAGTGGCGGCGCACGCGCGCGGAGGCG 1317
 QY 362 CAGCAGGCGAGCGAGGAGTGTGGGAGTACACCTGTGAGGCTGGGGCGCAGCGGCTC 421
 Db 1318 GAGGAGCGGCGGAGGAGCGCGCGGAGCGCGCGCTGCTGCGGATGGCGTGAGC-- 1375
 QY 422 TCCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTTTCAAGAGGAGCAGCTGGCACAC 481
 Db 1376 -CTGTGACGTGGAGGCTGAGGTGGCCAAAGCGGTGGAGGCTGTGCCAAGCGCGGTC 1434
 QY 482 AGG 484
 Db 1435 AAG 1437

RESULT 12

US-08-938-291A-3/c
 ; Sequence 3, Application US/08938291A
 ; Patent No. 6117673

; GENERAL INFORMATION:
 ; APPLICANT: Lev, Sima
 ; APPLICANT: Plowman, Gregory D.
 ; APPLICANT: Schlessinger, Joseph
 ; TITLE OF INVENTION: RDBG PROTEINS AND RELATED
 ; TITLE OF INVENTION: PRODUCTS AND METHODS
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; CITY: Suite 4700
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071-2066
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; MEDIUM TYPE: storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: FastSeq
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/938,291A
 ; FILING DATE: September 26, 1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/027,337
 ; FILING DATE: October 11, 1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warburg, Richard J.
 ; REGISTRATION NUMBER: 32,327
 ; REFERENCE/DOCKET NUMBER: 228/172
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (213) 489-1600
 ; TELEFAX: (213) 955-0440
 ; TELEX: 67-3510
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5020 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear


```
; Sequence 2, Application US/09010398
; Patent No. 5972615
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: Veltri, Robert W.
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,398
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Corder, Timothy S.
; REGISTRATION NUMBER: 38,414
; REFERENCE/DOCKET NUMBER: UROC:015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512-418-3000
; TELEFAX: 512-474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1709 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-010-398-2

Query Match      8.1%; Score 39.4; DB 2; Length 1709;
Best Local Similarity 47.1%; Pred. No. 0.25;
Matches 121; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy      6  GGTACAGGACGGGAAGCTGAGCTCCAGCTCAAAAGTACGCATGGAGGTCAGGGCT 65
Db      636  GGCTCAAGTATGAGATGAGCTGGCCCTGCCGCCAGGGCGTTGAGGCTGACATCAACGGCT 695

Qy      66  GCACACGAAGGCTGGTAGTCAGCAGCTGGGCAAGCAGATGCTGGGGAGTACAGCTGG 125
Db      696  TGGCCCGAGTCCTGGATGAGCTGACCTGGCCAGGACTGACCTGGAGATGCAGATCAGG 755

Qy      126  AGGCTGGGGCCAGAGTCTCTTCAATGCACATCAGAGCCCAAGGCGAGTGTGTTG 185
Db      756  GCCTGAATGAGGAGCTAGCTACCTGAAGAAAGAACCAAGAGGAGATGAAGGAGTTCA 815

Qy      186  CCAAGGAGCAGTTGGTGCATATAGGTGGGCTGAGGCGAGGGGCGAGTCCACACTGA 245
Db      816  GCAGCCAGCTGGCGGCGCGGCTAATGTGAGATGGACCGCAGCACCAGGGGTGTGGACCTGA 875

Qy      246  GCTGTGAGGTGGCCAG 262
Db      876  CCGTGTGCTGGCAGAG 892
```

Search completed: September 19, 2004, 22:12:35
Job time : 45.558 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 19, 2004, 19:28:00; Search time 251.035 Seconds
(without alignments)
9751.690 Million cell updates/sec

Title: US-10-077-130-4_COPY_3662_4146

Perfect score: 485

Sequence: 1 gacgtgtacagcagcgga.....ggagcagctggcacacagga 485

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/2/pubpna/FCR_NEW_PUB.seq.*
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/prodata/2/pubpna/FCRUS_PUBCOMB.seq.*
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 485 | 100.0 | 23907 | 14 | US-10-077-130-6 |
| 2 | 485 | 100.0 | 24120 | 14 | US-10-077-130-4 |
| 3 | 200.4 | 41.3 | 2768 | 16 | US-10-120-988-426 |
| 4 | 185.6 | 38.3 | 578 | 13 | US-10-027-632-272054 |
| 5 | 185.6 | 38.3 | 578 | 16 | US-10-027-632-272054 |
| 6 | 73.4 | 15.1 | 1645 | 9 | US-09-726-643-22 |
| 7 | 73.4 | 15.1 | 1645 | 14 | US-10-042-141-22 |
| 8 | 73.4 | 15.1 | 2170 | 16 | US-10-264-049-850 |
| 9 | 65.8 | 13.6 | 2534 | 16 | US-10-108-260A-1946 |
| 10 | 65.8 | 13.6 | 14061 | 16 | US-10-093-463-73 |
| 11 | 65.8 | 13.6 | 1409 | 16 | US-10-093-463-71 |
| 12 | 64.2 | 12.7 | 7564 | 17 | US-10-476-397-15 |
| 13 | 61.4 | 12.7 | 2254 | 9 | US-09-726-643-44 |
| 14 | 61.4 | 12.7 | 2254 | 14 | US-10-042-141-44 |

| | | | | | | |
|----|------|------|---------|----|---------------------|--------------------|
| 15 | 61 | 12.6 | 716 | 10 | US-09-622-846-117 | Sequence 117, App |
| 16 | 54.8 | 11.3 | 400 | 13 | US-10-085-783A-8373 | Sequence 8373, App |
| 17 | 54.8 | 11.3 | 400 | 16 | US-10-242-535A-8373 | Sequence 8373, App |
| 18 | 54.6 | 11.3 | 20565 | 10 | US-09-764-891-6045 | Sequence 6045, App |
| 19 | 54.6 | 11.3 | 20565 | 15 | US-10-091-438-270 | Sequence 270, App |
| 20 | 54.2 | 11.2 | 1266 | 10 | US-09-822-846-116 | Sequence 116, App |
| 21 | 53.6 | 11.1 | 2884 | 16 | US-10-094-749-1001 | Sequence 1001, App |
| 22 | 53.2 | 11.0 | 9591 | 10 | US-09-764-891-6043 | Sequence 6043, App |
| 23 | 53.2 | 11.0 | 9591 | 15 | US-10-091-438-268 | Sequence 268, App |
| 24 | 53.2 | 11.0 | 12415 | 10 | US-09-764-891-6044 | Sequence 6044, App |
| 25 | 53.2 | 11.0 | 12415 | 15 | US-10-091-438-269 | Sequence 269, App |
| 26 | 49.2 | 10.1 | 81940 | 9 | US-09-759-508B-1 | Sequence 1, Appli |
| 27 | 49.2 | 10.1 | 81940 | 10 | US-09-960-706-1092 | Sequence 1092, App |
| 28 | 49.2 | 10.1 | 81940 | 10 | US-09-873-319-724 | Sequence 724, App |
| 29 | 48.8 | 10.1 | 466 | 15 | US-10-029-386-22631 | Sequence 22631, A |
| 30 | 47.2 | 9.7 | 3575 | 17 | US-10-641-643-1395 | Sequence 1395, App |
| 31 | 47.2 | 9.7 | 3597 | 14 | US-10-098-841-84 | Sequence 84, Appl |
| 32 | 45 | 9.3 | 3935 | 16 | US-10-108-260A-534 | Sequence 534, App |
| 33 | 43.4 | 8.9 | 671 | 15 | US-10-184-644-346 | Sequence 346, App |
| 34 | 43.4 | 8.9 | 671 | 15 | US-10-184-634-346 | Sequence 346, App |
| 35 | 43.4 | 8.9 | 11916 | 15 | US-10-156-761-926 | Sequence 926, App |
| 36 | 43.4 | 8.9 | 11916 | 17 | US-10-204-862A-3 | Sequence 3, Appli |
| 37 | 43.4 | 8.9 | 16596 | 15 | US-10-156-761-930 | Sequence 930, App |
| 38 | 43.4 | 8.9 | 30690 | 17 | US-10-204-862A-1 | Sequence 1, Appli |
| 39 | 43.4 | 8.9 | 31422 | 17 | US-10-204-862A-2 | Sequence 2, Appli |
| 40 | 43.4 | 8.9 | 9025608 | 15 | US-10-156-761-1 | Sequence 1, Appli |
| 41 | 43.4 | 8.9 | 9025608 | 15 | US-10-156-761-1 | Sequence 1, Appli |
| 42 | 41 | 8.5 | 4916 | 13 | US-10-182-243-14 | Sequence 14, Appl |
| 43 | 41 | 8.5 | 5007 | 15 | US-10-307-619-3 | Sequence 3, Appli |
| 44 | 41 | 8.5 | 5207 | 9 | US-09-858-664A-1 | Sequence 1, Appli |
| 45 | 41 | 8.5 | 5207 | 13 | US-10-697-263-1 | Sequence 1, Appli |

ALIGNMENTS

RESULT 1

US-10-077-130-6
; Sequence 6, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses therefor
; FILE REFERENCE: MPI2001-047P/IRCP/1(M)
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6

| | | | | | | | |
|-----------------------|--------|--|----------|------------|----|--------|-------|
| Query Match | 100.0% | Score | 485 | DB | 14 | Length | 23907 |
| Best Local Similarity | 100.0% | Pred. No. | 1.5e-130 | Indels | 0 | Gaps | 0 |
| Matches | 485 | Conservative | 0 | Mismatches | 0 | Indels | 0 |
| QY | 1 | GACGTGTACAAAGGACGGGAAGAGCTGAGCTCCAGTCAAAAGTACGCATGGAGGTCAA | 60 | | | | |
| Db | 3591 | GACGTGTACAAAGGACGGGAAGAGCTGAGCTCCAGTCAAAAGTACGCATGGAGGTCAA | 3650 | | | | |
| QY | 61 | GGCTGTCACACAGAGGCTGTAGTCAGCAGTGGGCAAGCAGATCTGGGGATACAG | 120 | | | | |
| Db | 3651 | GGCTGTCACACAGAGGCTGTAGTCAGCAGTGGGCAAGCAGATCTGGGGATACAG | 3710 | | | | |
| QY | 121 | CTGCGAGGTGGGGCCACAGAGTCTCTTTCAATGCATCATCAGAGCCCAAGGCAGT | 180 | | | | |
| Db | 3711 | CTGCGAGGTGGGGCCACAGAGTCTCTTTCAATGCATCATCAGAGCCCAAGGCAGT | 3770 | | | | |

QY 181 GTTTCACAGGACAGTTGGTGCATAATGAGGTGCGGACTGAGGACAGGGGCCAGTGCAC 240
 |||||
 Db 3771 GTTTCACAGGACAGTTGGTGCATAATGAGGTGCGGACTGAGGACAGGGGCCAGTGCAC 3830
 |||||
 QY 241 ACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGTGGTACAAAGGATGGAGAA 300
 |||||
 Db 3831 ACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGTGGTACAAAGGATGGAGAA 3890
 |||||
 QY 301 GCTGAGCTCCAGTTCGAAAGTGGCATAGAGCTCGGGCTGCATGCGGACAGCTGGTGT 360
 |||||
 Db 3891 GCTGAGCTCCAGTTCGAAAGTGGCATAGAGCTCGGGCTGCATGCGGACAGCTGGTGT 3950
 |||||
 QY 361 GCAGCAGGACAGCCAGGACAGTCTGGGGAGTACACCTGTGAGCTGGGGGCCAGCGGCT 420
 |||||
 Db 3951 GCAGCAGGACAGCCAGGACAGTCTGGGGAGTACACCTGTGAGCTGGGGGCCAGCGGCT 4010
 |||||
 QY 421 CTCCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTTTGCAAAGGACAGCTGGCACA 480
 |||||
 Db 4011 CTCCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTTTGCAAAGGACAGCTGGCACA 4070
 |||||
 QY 481 CAGGA 485
 |||||
 Db 4071 CAGGA 4075

RESULT 2

US-10-077-130-4
 ; Sequence 4, Application US/10077130
 ; Publication No. US20020168742A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kapeller-Libermann, Rosana
 ; APPLICANT: Acton, Susan L.
 ; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
 ; FILE REFERENCE: MP12001-047P1RCP1(M)
 ; CURRENT APPLICATION NUMBER: US/10/077,130
 ; PRIORITY FILING DATE: 2002-02-15
 ; PRIOR FILING DATE: 2001-02-15
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 24120
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: 5'UTR
 ; LOCATION: (1)...(71)
 ; NAME/KEY: CDS
 ; LOCATION: (72)...(23978)
 ; NAME/KEY: 3'UTR
 ; LOCATION: (23979)...(24120)
 ; US-10-077-130-4

Query Match
 Best Local Similarity 100.0%; Score 485; DB 14; Length 24120;
 Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAGCTGTACAGGACCGGGAAGAGTGTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAA 60
 3662 GAGCTGTACAGGACCGGGAAGAGTGTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAA 3721
 |||||
 Db 61 GGCTGCACACGAGGCTGTGAGTGCAGAGGTGGGCAAGCAGATGCTGGGGAGTACAG 120
 3722 GGGCTGCACACGAGGCTGTGAGTGCAGAGGTGGGCAAGCAGATGCTGGGGAGTACAG 3781
 |||||
 QY 121 CTCGAGGCTGGGGGCCAGAGAGTCTCTTTCAACTGCACATCACAGAGCCCAAGGACGT 180
 3782 CTCGAGGCTGGGGGCCAGAGAGTCTCTTTCAACTGCACATCACAGAGCCCAAGGACGT 3841
 |||||
 QY 181 GTTTCACAGGACAGTTGGTGCATAATGAGGTGCGGACTGAGGACAGGGGCCAGTGCAC 240
 |||||

Db 3842 GTTTCACAGGACAGTTGGTGCATAATGAGGTGCGGACTGAGGACAGGGGCCAGTGCAC 3901
 |||||
 QY 241 ACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGTGGTACAAAGGATGGAGAA 300
 |||||
 Db 3902 ACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGTGGTACAAAGGATGGAGAA 3961
 |||||
 QY 301 GCTGAGCTCCAGTTCGAAAGTGGCATAGAGCTCGGGCTGCATGCGGACAGCTGGTGT 360
 |||||
 Db 3962 GCTGAGCTCCAGTTCGAAAGTGGCATAGAGCTCGGGCTGCATGCGGACAGCTGGTGT 4021
 |||||
 QY 361 GCAGCAGGACAGCCAGGACAGTCTGGGGAGTACACCTGTGAGCTGGGGGCCAGCGGCT 420
 |||||
 Db 4022 GCAGCAGGACAGCCAGGACAGTCTGGGGAGTACACCTGTGAGCTGGGGGCCAGCGGCT 4081
 |||||
 QY 421 CTCCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTTTGCAAAGGACAGCTGGCACA 480
 |||||
 Db 4082 CTCCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTTTGCAAAGGACAGCTGGCACA 4141
 |||||
 QY 481 CAGGA 485
 |||||
 Db 4142 CAGGA 4146

RESULT 3

US-10-120-988-426
 ; Sequence 426, Application US/10120988
 ; Publication No. US20030219745A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Goodrich, Ryle
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Ren, Feiyan
 ; APPLICANT: Wang, Dunrui
 ; APPLICANT: Drmanac, Radoje T.
 ; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
 ; FILE REFERENCE: 802CON
 ; CURRENT APPLICATION NUMBER: US/10/120,988
 ; PRIORITY FILING DATE: 2002-04-11
 ; PRIOR FILING DATE: 2001-01-30
 ; NUMBER OF SEQ ID NOS: 441
 ; SOFTWARE: pt_FL_genes Version 2.0
 ; SEQ ID NO 426
 ; LENGTH: 2768
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (72)...(2738)
 ; US-10-120-988-426

Query Match
 Best Local Similarity 41.3%; Score 200.4; DB 16; Length 2768;
 Matches 286; Conservative 0; Mismatches 121; Indels 3; Gaps 1;
 QY 5 TGGTACAGGACGGGAAGAGTGTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAAGGCG 64
 2349 TGGTACAGGATGGCATGGAGCTGGGCACTCCGGTGTGAGCGCTCTTTCAGGAGGATGTG 2408
 |||||
 Db 65 TGCACAG--AAGGCTGTAGTGCAGCAGTGGGCAAGCAGATGCTGGGAGTATACAG 121
 2409 GGGACGGGCAACCGGCTGGTGGCAGCCACAGTCAACAGGAGGATGAAGCACCCTACTCC 2468
 |||||
 QY 122 TGCAGGCTGGGGGCCAGAGAGTCTCCTTTCACTGCACATCACAGAGCCCAAGGAGTG 181
 2469 TGCAGGCTGGGGGCCAGAGAGTCTCCTTTCACTGCAGCTTCCGGCTCTCTGAGCCCAAGGAGTG 2528
 |||||
 QY 182 TTTTCCAGAGAGCAGTTGGTGCATAATGAGGTGCGGACTGAGGACAGGGGCCAGTGCACA 241
 2529 TTTTCCAGAGAGCAGTTGGTGCATAATGAGGTGCGGCTGAGGTGGGGGCCAGGCGCACG 2588
 |||||
 QY 242 CTGAGCTGTGAGGTGGGCCAGGCGCCAGAGAGGTGACGCTGAGTGTGTTACAAGGATGGGAAG 301
 |||||

Db 2589 CTGAGCTGTGAGGTGGCCAGGACCATGATGAGGTGACGTGGTACAGGACGGGAAG 2648
Qy 302 CTGAGCTCCAGTTCGAAGTGGCATAGAGCTGCGGCTGCATGGCGCAGCTGGTGGT 361
Db 2649 TTGAGCTCCAGCTCGAAGTGCAGTGGAGGCTGTGGGCTGTATGCGAGGCTGGTGGT 2708
Qy 362 CAGCAGGAGCCAGGACGATGCTGGGAGGTACACCTGTGAGGTGGGG 411
Db 2709 CAGCAGGTGGCCAGGACAGCTCCGATAGTACAGCTGTGAAGCCAGGTG 2758

RESULT 4
US-10-027-632-272054
; Sequence 272054, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272054
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-272054

Query Match 38.3%; Score 185.6; DB 13; Length 578;
Best Local Similarity 87.5%; Pred. No. 6.2e-44;
Matches 203; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
Qy 163 CACAGAGCCCAAGGACAGTGTTCCTCAAGGAGCAGTGTGCTATATGAGTGGCGACTGA 222
Db 347 CCAGAGCCCAAGTGGTGTTCCTCAAGGAGCAGCAGCAGGAGGTGCAGGCTGA 406
Qy 223 GGCAGGGCCAGTGCACACTGAGTGTGAGTGGCCAGGCCAGCAGAGAGGTGACGTG 282
Db 407 GCGGGGGCCAGTGCACCTGAGTGTGAGTGGCCAGGCCAGGCCAGCAGAGAGGTGACGTG 466
Qy 283 GTACAGGATGGAGAGCTCAGTCCAGTTCGAAAGTGGCATAGAGGCTGCGGGCTG 342
Db 467 GTACAGGATGGAGAGCTCAGTTCGAAAGTGGCATAGGCTGCGGGCTG 526
Qy 343 CATGCGGACAGCTGGTGGTGCAGCAGGAGCCAGGACAGATGCTGGGGAGTAC 394
Db 527 CACACGAGGCTGGTGGTGCAGCAGGCGGGCCAGGACAGAGCGGGGAGTAC 578

RESULT 5
US-10-027-632-272054
; Sequence 272054, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 272054
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-272054
Query Match 38.3%; Score 185.6; DB 16; Length 578;
Best Local Similarity 87.5%; Pred. No. 6.2e-44;
Matches 203; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
Qy 163 CACAGAGCCCAAGGACAGTGTTCCTCAAGGAGCAGTGTGCTATATGAGTGGCGACTGA 222
Db 347 CCAGAGCCCAAGTGGTGTTCCTCAAGGAGCAGCAGCAGGAGGTGCAGGCTGA 406
Qy 223 GGCAGGGCCAGTGCACACTGAGTGTGAGTGGCCAGGCCAGCAGAGAGGTGACGTG 282
Db 407 GCGGGGGCCAGTGCACCTGAGTGTGAGTGGCCAGGCCAGGCCAGCAGAGAGGTGACGTG 466
Qy 283 GTACAGGATGGAGAGCTCAGTCCAGTTCGAAAGTGGCATAGAGGCTGCGGGCTG 342
Db 467 GTACAGGATGGAGAGCTCAGTTCGAAAGTGGCATAGGCTGCGGGCTG 526
Qy 343 CATGCGGACAGCTGGTGGTGCAGCAGGAGCCAGGACAGATGCTGGGGAGTAC 394
Db 527 CACACGAGGCTGGTGGTGCAGCAGGCGGGCCAGGACAGAGCGGGGAGTAC 578

RESULT 6
US-09-726-643-22
; Sequence 22, Application US/09726643
; Patent No. US200202849A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-726-643-22

Query Match 15.1%; Score 73.4; DB 9; Length 1645;
Best Local Similarity 49.4%; Pred. No. 2.4e-11;
Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 5 TGGTACAAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGGTCAAGGGC 64
 Db 905 TGGTACAAGGACGGGACGAGAGTGGAGGAGTACTTCGTGGTCTGGAGATGAGGGG 964
 QY 65 TGCACACGAAGGCTGGTGTAGTCAGCAGAGTGGGCAAGAGATGCTGGGAGTACAGCTGC 124
 Db 965 CCCCATCGCCGCTGGTGTGTCGCCGCCACCCATCCCTCAGACGGGGCGAGTTTCAGTGC 1024
 QY 125 GAGGCTGGGGCCAGAGAGTCTCCCTTTCAACTGCACATCACAGAGCCCAAGCGAGTGT 184
 Db 1025 GTCCGTGGAGATGAGTGTGCTACTTCACTGTACCATCACAGAGCTCTCCTGTGATC 1084
 QY 185 GCCAAGGACAGTGGTGTGATATGAGTGGGATGAGGAGGAGGAGGAGGAGGAGGAGG 244
 Db 1085 GT---GTATCCAGCGGCAAGGTGTATGTGGCAGCGCTCGGCTGGAGCGTGTGTGCTG 1141
 QY 245 AGCTGTGAGGTGGCCAGGAGTCTGCTGGGAGTACACCTGTGAGGCTGGGGGCGAGCGGCTCTCC 424
 Db 1262 GCTGTCCAGCTCGAGGACTCCGCGGAGTACTTGTGTGAAATTGACGATGAGTCGGCTCTCC 1321
 QY 425 TTCCACCTGGATGTTTCAGAGCC 447
 Db 1322 TTCACTGTCCAGCTCACAGATC 1344
 RESULT 7
 US-10-042-141-22
 ; Sequence 22, Application US/10042141
 ; Publication No. US20020183503A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: 26 Human secreted proteins
 ; FILE REFERENCE: P2040P1
 ; CURRENT APPLICATION NUMBER: US/10/042,141
 ; CURRENT FILING DATE: 2002-01-11
 ; PRIOR APPLICATION NUMBER: 09/726,643
 ; PRIOR FILING DATE: 2000-12-01
 ; PRIOR APPLICATION NUMBER: PCT/US00/15187
 ; PRIOR FILING DATE: 2000-06-02
 ; PRIOR APPLICATION NUMBER: 60/137,725
 ; PRIOR FILING DATE: 1999-06-07
 ; NUMBER OF SEQ ID NOS: 190
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 22
 ; LENGTH: 1645
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-042-141-22

Query Match 15.1%; Score 73.4; DB 14; Length 1645;
 Best Local Similarity 49.4%; Pred. No. 2.4e-11;
 Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;
 QY 5 TGGTACAAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGGTCAAGGGC 64
 Db 905 TGGTACAAGGACGGGACGAGAGTGGAGGAGTACTTCGTGGTCTGGAGATGAGGGG 964
 QY 65 TGCACACGAAGGCTGGTGTAGTCAGCAGAGTGGGCAAGAGATGCTGGGAGTACAGCTGC 124
 Db 965 CCCCATCGCCGCTGGTGTGTCGCCGCCACCCATCCCTCAGACGGGGCGAGTTTCAGTGC 1024
 QY 125 GAGGCTGGGGCCAGAGAGTCTCCCTTTCAACTGCACATCACAGAGCCCAAGCGAGTGT 184
 Db 1025 GTCCGTGGAGATGAGTGTGCTACTTCACTGTACCATCACAGAGCTCTCCTGTGATC 1084

QY 185 GCCAAGGACAGTGGTGTGATATGAGGTGGGAGCTGAGGACGAGGCGCAGTCCACACTG 244
 Db 1085 GT---GTATCCAGCGGCAAGGTGTATGTGGCAGCGCTCGGCTGGAGCGTGTGTGCTG 1141
 QY 245 AGCTGTGAGGTGGCCAGGAGTCTGCTGGGAGTACACCTGTGAGGCTGGGGGCGAGCGGCTCTCC 424
 Db 1142 ACCTGTGAGCTATGCGCGGCTCGGCGGAGTGGCTGGACCAAGGATGGAGAGGAGTGT 1201
 QY 305 AGCTCCAGTTCGAAAGTGGCATAGAGGCTGGGAGTGCATCGCGGAGTGCATCGCGAGTGTGTGTCAG 364
 Db 1202 GTGGAGAGCCCGCGCTGCTCTGCAAGAGGAGACACTGTCCCGCGCTGGTGTGCTGCC 1261
 QY 365 CAGGACGGCCAGGACAGATGCTGGGAGTACACCTGTGAGGCTGGGGGCGAGCGGCTCTCC 424
 Db 1262 GCTGTCCAGCTCGAGGACTCCGCGGAGTACTTGTGTGAAATTGACGATGAGTCGGCTCTCC 1321
 QY 425 TTCCACCTGGATGTTTCAGAGCC 447
 Db 1322 TTCACTGTCCAGCTCACAGATC 1344
 RESULT 8
 US-10-264-049-850
 ; Sequence 850, Application US/10264049
 ; Publication No. US20040005579A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birse et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P4133P1
 ; CURRENT APPLICATION NUMBER: US/10/264,049
 ; CURRENT FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/18569
 ; PRIOR FILING DATE: 2001-06-07
 ; PRIOR APPLICATION NUMBER: US 60/209,467
 ; PRIOR FILING DATE: 2000-06-07
 ; NUMBER OF SEQ ID NOS: 4360
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 850
 ; LENGTH: 2170
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-264-049-850

Query Match 15.1%; Score 73.4; DB 16; Length 2170;
 Best Local Similarity 49.4%; Pred. No. 2.5e-11;
 Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;
 QY 5 TGGTACAAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGGTCAAGGGC 64
 Db 1396 TGGTACAAGGACGGGACGAGAGTGGAGGAGTACTTCGTGGTCTGGAGATGAGGGG 1455
 QY 65 TGCACACGAAGGCTGGTGTAGTCAGCAGAGTGGGCAAGAGATGCTGGGAGTACAGTGC 124
 Db 1456 CCCCATCGCCGCTGGTGTGTCGCCGCCACCCAGCGCTCAGACGGGGCGAGTTTCAGTGC 1515
 QY 125 GAGGCTGGGGCCAGAGAGTCTCCCTTTCAACTGCACATCACAGAGCCCAAGCGAGTGT 184
 Db 1516 GTCCGTGGAGATGAGTGTGCTACTTCACTGTACCATCACAGAGCTCTCCTGTGATC 1575
 QY 185 GCCAAGGACAGTGGTGTGATATGAGGTGGGATGAGGAGGAGGAGGAGGAGGAGGAG 244
 Db 1576 GT---GTATCCAGCGGCAAGGTGTATGTGGCAGCGCTCGGCTGGAGCGTGTGTGCTG 1632
 QY 245 AGCTGTGAGGTGGCCAGGAGTGCATAGAGGCTGGGAGTGCATGAGGATGGGAGAGCTG 304
 Db 1633 ACCTGTGAGCTATGCGCGCTCGGCGGAGTGGCTGGACCAAGGATGGAGAGGAGTGT 1692
 QY 305 AGCTCCAGTTCGAAAGTGGCATAGAGGCTGGGGCTGCATGCGGAGTGTGTGTCAG 364
 Db 1693 GTGGAGAGCCCGCGCTGCTCCTSCAAGAGGAGACACTGTCCGCGCGCTGGTGTGCC 1752
 QY 365 CAGGACGGCCAGGAGATGCTGGGAGTACACTGTGAGGCTGGGGGCGAGCGGCTCTCC 424

Db 1753 GCTGTCCAGCTCGAGGACTCGGCGACTTGTGTGAATTGACGATGAGTCGGGCTCC 1812
QY 425 TTCCACCTGGATGTTTCAGAGCC 447
Db 1813 TTCACTGTCCCGTCACAGATC 1835

RESULT 9

US-10-108-260A-1946
; Sequence 1946, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1946
; LENGTH: 2534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1946

Query Match 13.6%; Score 65.8; DB 16; Length 2534;
Best Local Similarity 51.9%; Pred. No. 4.1e-09;
Matches 148; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 194 CAGTTGGTGCAATATGAGTGGGACCTGAGGAGGCGGCGGACCTGCCACACTGAGCTGTGAG 253
Db 256 CTGAAGGCGCTGGATGACCTGTCCGACAGAGGCGGCGGACCTGGGCGCTGCAGTGTGAA 315
QY 254 GTGGCCCGAGCCAGACAGAGTGACGTGGTACAGGATGGAGAGAGCTGAGCTCCAGT 313
Db 316 GTCTGTGACCGCGAGGCGGCGGCTGTGGTGTGGCGCAAGATGGCTGCGAGCTGGGCGGCGG 375
QY 314 TCGAAGTGGCATAGAGGCTGCGGGCTGCAATGCGGCGAGCTGGTGTGACGAGGCGGCGG 373
Db 376 GACAAGTATGACTTCTCTGCACAGGCGGCGGCGGCGGCTGCTGGTGCATGACGTGAGC 435
QY 374 CAGGCGAGATGTGGGAGTACACCTGTGAGGCTGGGGGCGGAGCGGCTCTCCTTCGACCTG 433
Db 436 CTTGAAGACGCGGCGGCTGTACACCTGCCACGCTGGGCTCCGAGGAGACCGCGGCGGCGG 495
QY 434 GATGTTTCAAGCCCAAGCGGTGTTTCAAGAGGAGAGCTGGCA 478
Db 496 CCGGTGCAGATCTGCACGTGGGCGATCCAAAGAGGCTGAAGACA 540

RESULT 10

US-10-093-463-73
; Sequence 73, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Basha
; APPLICANT: Vernet, Corine

; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1el Antibodies that bind to Antigenic Polypept
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 73
; LENGTH: 14061
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15)..(14039)
US-10-093-463-73

Query Match 13.6%; Score 65.8; DB 16; Length 14061;
Best Local Similarity 51.9%; Pred. No. 5.1e-09;
Matches 148; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 194 CAGTTGGTGCAATATGAGTGGGACCTGAGGAGGCGGCGGACCTGCCACACTGAGCTGTGAG 253
Db 180 CTGAAGGCGCTGGATGACCTGTCCGACAGAGGCGGCGGACCTGGGCGCTGCAGTGTGAA 239
QY 254 GTGGCCCGAGCCAGACAGAGTGACGTGGTGAAGAGAGAGCTGCCAGTCCAGTCCAGT 313

Db 240 GTCTCTGACCCGAGGCGCCATGTGGTGGCGCAAGATGGCGTGCAGCTGGGCCCACT 299
QY 314 TCGAAGTGGCATAGAGCTCGGGCTGCATCGCGCAGCTGGTGGTGCAGAGCAGGC 373
Db 300 GACAAGTATGACTTCTCTGCACACGCGCGGCGCAGCGCGGGGCTCGTGGTGCATGAGCTGAGC 359
QY 374 CAGGCAGATGCTGGGCGAGTACACCTGTGAGGCTGGGGCCAGCGGCTCTCTTCCACTG 433
Db 360 CTGAGAGCGCGGCTGTACACCTGCACGTGGGCTCGAGGAGACCGGGCCCGGCTC 419
QY 434 GATGTTTCAGAGCCCAAGCGGTGTTTGCAGAGGAGCAGCTGGCA 478
Db 420 CCGGTGCACGATCTGCAGCTGGGCATCACCAGAGGCTGAAAGACA 464

RESULT 11

US-10-093-463-71

; Sequence 71, Application US:10093463

; Publication No. US20030208039A1

; GENERAL INFORMATION:

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Shenoy, Suresh

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Gusev, Vladimir

; APPLICANT: Pochart, Pascal

; APPLICANT: Zhong, Mei

; APPLICANT: Rastelli, Luca

; APPLICANT: Mezes, Peter

; APPLICANT: Smithson, Glennda

; APPLICANT: Guo, Xiaojia

; APPLICANT: Gerlach, Valerie

; APPLICANT: Casman, Stacie

; APPLICANT: Boldog, Ferenc

; APPLICANT: Li, Li

; APPLICANT: Zethusen, Bryan

; APPLICANT: Tchernev, Velizar

; APPLICANT: Gangolli, Esha

; APPLICANT: Vernet, Corine

; APPLICANT: Pena, Carol

; APPLICANT: Burgess, Catherine

; APPLICANT: Liu, Xiaohong

; APPLICANT: Spytek, Kimberly

; APPLICANT: Gorman, Linda

; APPLICANT: Spaderna, Steven

; APPLICANT: Voss, Edward

; APPLICANT: Malyankar, Uriel

; APPLICANT: Anderson, David

; APPLICANT: Patturajan, Meera

; APPLICANT: Miller, Charles

; APPLICANT: Taupier, Raymond J. Jr.

; TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypepti

; FILE OF INVENTION: Encoding the Antigens, and Methods of Use.

; FILE REFERENCE: 21402-290A (Cura 590AT)

; CURRENT APPLICATION NUMBER: US/10/093,463

; CURRENT FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 60/283,675

; PRIOR FILING DATE: 2001-04-14

; PRIOR APPLICATION NUMBER: 60/338,092

; PRIOR FILING DATE: 2001-12-03

; PRIOR APPLICATION NUMBER: 60/274,281

; PRIOR FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: 60/274,101

; PRIOR FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: 60/325,681

; PRIOR FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/304,354

; PRIOR FILING DATE: 2001-07-10

; PRIOR APPLICATION NUMBER: 60/279,995

; PRIOR FILING DATE: 2001-03-30

; PRIOR APPLICATION NUMBER: 60/294,899

; PRIOR FILING DATE: 2001-05-31

; PRIOR APPLICATION NUMBER: 60/287,424

; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
; LENGTH: 14109
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15)..(14088)
US-10-093-463-71

Query Match 13.6%; Score 65.8; DB 16; Length 14109;

Best Local Similarity 51.9%; Pred. No. 5.1e-09;

Matches 148; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 194 CAGTTGGTGCATATGAGGTGGGACTGAGGCTGAGGCGGCGCAGTGCACACTGAGTGTGAG 253
Db 180 CTGAAGGCGCTGGATGACCTGTCCGACAGAGGCGCGGCACCCCTGGCCCTCGAGTGTGAA 239
QY 254 GTGGCCAGGCGCCACAGACAGAGGTGACGTGTGTACAGGATGGGAAAGCTGAGTCCAGT 313
Db 240 GTCTCTGACCCCGAGGCCCATGTGTGTGTGGCGCAAGATGCGTGCAGTGGCCCCAGT 299
QY 314 TCGAAGTGGCATAGAGGCTGCGGGCTGCATGCGGCGAGCTGGTGTGTGCACAGGCGAGC 373
Db 300 GACAAGTATGACTTCTCTGCACACGCGCGGCGCAGCGGGGCTCGTGGTGCATGAGCTGAGC 359
QY 374 CAGGCAGATGCTGGGGAGTACACCTGTGAGGCTGGGGCCAGCGGCTCTCTTCCACTG 433
Db 360 CCTGAAGACGCGCGGCTGTACACCTGCCACGTGGGCTCCGAGGAGAGCCCGGGCCCGGCTC 419
QY 434 GATGTTTCAGAGCCCAAGCGCGTGTGTTGCAAGGAGCAGCTGGCA 478
Db 420 CCGGTGCACGATCTGCACGTGGGCATCACCAGAGGCTGAAAGACA 464

RESULT 12

US-10-476-397-15

; Sequence 15, Application US:10476397

; Publication No. US20040115687A1

; GENERAL INFORMATION:

; APPLICANT: YUE, Henry

; APPLICANT: LEE, Ernestine A.

; APPLICANT: DUGGAN, Brendan M.

; APPLICANT: THANGAVELU, Kavitha

; APPLICANT: HONCHELL, Cynthia D.

; APPLICANT: DING, Li

; APPLICANT: JACKSON, Jennifer L.

; APPLICANT: BAUGHN, Mariah R.

; APPLICANT: KALLICK, Deborah A.

; APPLICANT: LEE, Sally

; APPLICANT: WARREN, Bridget A.

; APPLICANT: XU, Yuming

; APPLICANT: TRAN, Uyen K.

/ APPLICANT: LAL, Preeti G.
 / APPLICANT: THORNTON, Michael B.
 / APPLICANT: HAFALIA, April J.A.
 / APPLICANT: YAO, Monique G.
 / APPLICANT: NGUYEN, Dannel B.
 / APPLICANT: GANDHI, Ameena R.
 / APPLICANT: KHAN, Farrah A.
 / APPLICANT: CHAWLA, Narinder K.
 / APPLICANT: GRIFFIN, Jennifer A.
 / APPLICANT: CHINN, Anna M.
 / APPLICANT: ELLIOTT, Vicki S.
 / APPLICANT: RAMKUMAR, Javalaxmi
 / APPLICANT: ARVIZU, Chandra S.
 / APPLICANT: FORSTHE, Ian J.
 / TITLE OF INVENTION: CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS
 / FILE REFERENCE: PF-0968 USN
 / CURRENT APPLICATION NUMBER: US/10/476,397
 / CURRENT FILING DATE: 2003-10-31
 / PRIOR APPLICATION NUMBER: PCT/US02/13874
 / PRIOR FILING DATE: 2002-05-01
 / PRIOR APPLICATION NUMBER: US 60/288,290
 / PRIOR FILING DATE: 2001-05-02
 / PRIOR APPLICATION NUMBER: US 60/292,468
 / PRIOR FILING DATE: 2001-05-21
 / PRIOR APPLICATION NUMBER: US 60/298,616
 / PRIOR FILING DATE: 2001-06-15
 / PRIOR APPLICATION NUMBER: US 60/301,672
 / PRIOR FILING DATE: 2001-06-28
 / PRIOR APPLICATION NUMBER: US 60/345,008
 / PRIOR FILING DATE: 2002-01-04
 / NUMBER OF SEQ ID NOS: 22
 / SOFTWARE: PERL Program
 / SEQ ID NO 15
 / LENGTH: 7564
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 / NAME/KEY: misc feature
 / OTHER INFORMATION: Incyte ID No: 7326129CB1
 / US-10-476-397-15.

Query Match 13.2%; Score 64.2; DB 17; Length 7564;
 Best Local Similarity 51.6%; Pred. No. 1.4e-08;
 Matches 147; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

| | | | |
|----|------|--|------|
| Qy | 194 | CAGTTGTGTCATATGAGTGGCGACCTGAGCGAGGGCCAGTGCACACTGAGCTGTGAG | 253 |
| Db | 2065 | CTGAAGCGCTGGATGCTCCGACAGAGGCGCGCCACCCCTGGCCCTGCAGTGTGAA | 2124 |
| Qy | 254 | GTGCCCCAGGCCCCAGACAGAGGTGACCTGTGTAAGAAGATGGGAAGAGCTGAGTCCAGT | 313 |
| Db | 2125 | GTCTCTGACCCCGAGGCCATGTGTGTGGCGCAAGATGGCGTGCAGCTGGGCCCCCAGT | 2184 |
| Qy | 314 | TCGAAAGTGGCATAGAGGCTGGCGGCTGCATGCGGCGAGCTGGTGGTGCAGCAGGCAGGC | 373 |
| Db | 2185 | GACAAGTATGACTTCTTGCAACCGCGGACACGCGGGGCTCTGTGTGCATGACGTGAGC | 2244 |
| Qy | 374 | CAGGCAGATGCTGGGGAGTACACTGTGAGGCTGGGGGCCAGCGGCTCTCTCTCCACCTG | 433 |
| Db | 2245 | CCTGAAGACGCGCGCTGTACACCTGCCACATGGGCTCCGAGGAGACCCCGGCCCGGTC | 2304 |
| Qy | 434 | GATGTTTCAGACCCAGCGCGTGTTCGAAAGGAGCAGCTGGCA | 478 |
| Db | 2305 | CGCGTGCAACGATCTGCAGTGGGCATACCAAGAGGCTGAAGACA | 2349 |

RESULT 13
 US-09-726-643-44
 ; Sequence 44, Application US/09726643
 ; Patent No. US2002028449A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: 26 Human secreted proteins

```
; ORGANISM: Homo sapiens
US-10-042-141-44

Query Match      12.7%; Score 61.4; DB 14; Length 2254;
Best Local Similarity 49.2%; Pred. No. 7.6e-08;
Matches 218; Conservative 0; Mismatches 221; Indels 4; Gaps 2;

QY 5 TGGTACAAGCAGCGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGGTCAAGGGC 64
Db 1400 TGGTACAAGCAGCGGAGAGCTGAGGAGAGTACTTCGTGGTCTCGAGAAATCAGGGG 1459
QY 65 TCCACACGAGAGCTGGTAGTCAGAGGTGGGCAAGCAGATGCTGGGAGATCAGCTGC 124
Db 1460 CCCCATCGCGCTCGTGGTGGTCCGCCACCACCCAGCCCTCAGACGGGGGCGAGTTTCAGTGC 1519
QY 125 GAGGCTGGGGGCCAGAGAGTCTCTCTTTCAACTGCACATCACAGAGCCCAAGGCCAGTGT 184
Db 1520 GTCTGCTGGAGATGAGTGTGCTTACTGTCTACCATCACAGAGTCTCTCTGTGGATC 1579
QY 185 GCCAAGGAGCAGTTGGTGGATATAGAGTGGGACTGAGGAGGGGCCAGTGCACACTG 244
Db 1580 GT---GTATCCAGCGCAAGGTGTATGTGGCAGCGTGGCCCTGGAGCGTGTGGTGTG 1636
QY 245 AGCTGTGAGGTGGCCAGCGCCAGAGAGGTGAGTGTGTTACAAAGTGGGAAGCAAGCTG 304
Db 1637 ACCTGTGAGTATGCGCGCCCTGGGAGAGGTGGCTGGACCAAGGATGGAGAGAGGTG 1696
QY 305 AGCTCCAGTTTGAAGTGGCATAGAGCTCGGGCTGCATGCGGCGAGCTGGTGGTCCAG 364
Db 1697 GTGGAGAGCCCG-CGCTGCTCTCTGTCAGAAAGAGACACTGTCCGCCCTGGTGTGCTGCC 1755
QY 365 CAGGCGAGCCAGGCGAGATGCTGGGGAGTACACCTGTGAGGCTGGGCGCAGCGCTCTCC 424
Db 1756 GCTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAATTGACGATGAGTGGCGCTCC 1815
QY 425 TTCCACCTGGAGTTTCAGAGCC 447
Db 1816 TTCCTGTACCCCTCACAGAGTC 1838

Search completed: September 20, 2004, 02:51:55
Job time : 261.035 secs
```

```
RESULT 15
US-09-822-846-117
; Sequence 117, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; PRIORITY FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/135,605
; PRIORITY FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 117
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 19, 2004, 15:07:04 ; Search time 1414.09 Seconds
(without alignments)
10242.013 Million cell updates/sec

Title: US-10-077-130-4_COPY_3662_4146
Perfect score: 485
Sequence: 1 gagctggtacaaaggacggga.....ggagcagctggcacacagga 485

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rod:*
 - 26: em_gss_phg:*
 - 27: em_gss_vrl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| C 1 | 342.2 | 70.6 | 632 | 14 | CF180572 | CF180572 815696 MA |
| C 2 | 314.8 | 64.9 | 621 | 14 | CF179218 | CF179218 813315 MA |
| C 3 | 216 | 44.5 | 1142 | 12 | B680164 | B680164 60262828 |
| C 4 | 212.8 | 43.9 | 401 | 10 | AW898791 | AW898791 CMO-NN007 |

| | | | | | | |
|------|-------|------|------|----|----------|--------------------|
| 5 | 207.6 | 42.8 | 522 | 28 | AQ595824 | AQ595824 HS_2132 B |
| 6 | 207.2 | 42.7 | 591 | 13 | BU439885 | BU439885 604144795 |
| 7 | 194.8 | 40.2 | 658 | 28 | AZ383233 | AZ383233 1M0140117 |
| C 8 | 193.8 | 40.0 | 574 | 9 | AI595154 | AI595154 ml39h10.Y |
| 9 | 180.4 | 37.2 | 786 | 14 | CD099595 | CD099595 AGENCOURT |
| 10 | 161.2 | 33.2 | 588 | 13 | BU438399 | BU438399 604145537 |
| C 11 | 142.4 | 29.4 | 373 | 10 | BE159678 | BE159678 MRO-HT040 |
| C 12 | 140.6 | 29.0 | 343 | 9 | AA061264 | AA061264 ml39h10.Y |
| 13 | 132.4 | 27.3 | 631 | 12 | BG793386 | BG793386 UTSW_SMI1 |
| 14 | 132.2 | 27.3 | 397 | 28 | B2837763 | B2837763 CH240_248 |
| 15 | 127.8 | 26.4 | 1192 | 28 | CC247513 | CC247513 PM2-127 |
| 16 | 112.6 | 23.2 | 179 | 12 | BG989628 | BG989628 PM2-HT117 |
| 17 | 112.4 | 23.2 | 472 | 29 | CB495069 | CB495069 tigr-gss- |
| C 18 | 85.8 | 17.7 | 724 | 29 | CNS03WVR | AL264240 Tetraodon |
| C 19 | 84.4 | 17.4 | 875 | 29 | CNS03WVR | AL264239 Tetraodon |
| 20 | 75.6 | 15.6 | 470 | 10 | BF087837 | BF087837 IL3-HT061 |
| 21 | 74 | 15.3 | 612 | 13 | BU947303 | BU947303 io46c03.Y |
| C 22 | 73.4 | 15.1 | 525 | 10 | BF087734 | BF087734 IL3-HT061 |
| 23 | 73.4 | 15.1 | 886 | 13 | BU538739 | BU538739 AGENCOURT |
| C 24 | 73.4 | 15.1 | 982 | 13 | BQ877554 | BQ877554 AGENCOURT |
| 25 | 73.4 | 15.1 | 1085 | 12 | BM546177 | BM546177 AGENCOURT |
| 26 | 73.4 | 15.1 | 1098 | 12 | BM543727 | BM543727 AGENCOURT |
| 27 | 73.4 | 15.1 | 1201 | 13 | BM377136 | BM377136 |
| 28 | 73.4 | 15.1 | 2066 | 29 | AY399870 | AY399870 Homo sapi |
| 29 | 72.4 | 14.9 | 505 | 10 | BF087860 | BF087860 IL3-HT061 |
| C 30 | 72 | 14.8 | 805 | 13 | BM381923 | BM381923 |
| 31 | 71.8 | 14.8 | 575 | 13 | BU072543 | BU072543 im44e07.Y |
| 32 | 71.6 | 14.8 | 569 | 12 | BM254228 | BM254228 515922 MA |
| C 33 | 71.4 | 14.7 | 982 | 13 | BM343487 | BM343487 |
| C 34 | 71 | 14.6 | 533 | 10 | AW995362 | AW995362 QVO-BN004 |
| 35 | 70.4 | 14.5 | 4130 | 11 | AK048144 | AK048144 Mus muscu |
| 36 | 70.4 | 14.5 | 5558 | 11 | AK029863 | AK029863 Mus muscu |
| 37 | 67.8 | 14.0 | 555 | 9 | AU119815 | AU119815 |
| 38 | 67.8 | 14.0 | 724 | 9 | AU141131 | AU141131 |
| C 39 | 67.8 | 14.0 | 770 | 12 | BG819323 | BG819323 602781512 |
| 40 | 67.8 | 14.0 | 1201 | 13 | BM377135 | BM377135 |
| 41 | 67 | 13.8 | 649 | 14 | CF262757 | CF262757 AUA_Ipfit |
| 42 | 66 | 13.6 | 743 | 10 | BE314256 | BE314256 601153138 |
| 43 | 66 | 13.6 | 835 | 10 | BE260056 | BE260056 601150922 |
| 44 | 66 | 13.6 | 949 | 13 | BQ645083 | BQ645083 AGENCOURT |
| 45 | 66 | 13.6 | 949 | 13 | BQ839538 | BQ839538 AGENCOURT |

ALIGNMENTS

RESULT 1
CF180572/c
LOCUS 815696 MARC 3PIG Sus scrofa cDNA 3', mRNA sequence.
DEFINITION CF180572.1 GI:33292348
ACCESSION
VERSION
KEYWORDS
SOURCE Sus scrofa (pig)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 632)
AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J., Wise,T.A., Noneman,D.J., Wray,J.E. and Keele,J.W.
TITLE A second set of porcine ESTs from a pooled-tissue normalized library
JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross_match v0.990329.
Plate: SRG8009 row: O column: 4

```

Seq primer: TAGAAGGCACAGCTCGAGG.
Location/Qualifiers
1..632
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3P1G"
/note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

ORIGIN
Query Match 70.6%; Score 342.2; DB 14; Length 632;
Best Local Similarity 81.8%; Pred. No. 4e-65;
Matches 395; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1 GACGTGGTACAGGACGGGAGAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAA 60
Db 512 GACTTGGTACAGGATGGGAGAGAGCTGAGTTCAGAGCTCAAAAGTGGGTGGAGGCCAA 453

QY 61 GGGGTGCACACGAAGGCTGTGTGTCAGCAGAGGTGGGCAAGCAGATGCTGGGAGTACAG 120
Db 452 GGGGTGCACACCGAGGCTGTGTGTCAGCAAGTGGGCAAGCAGATGCTGGGAGTACAG 393

QY 121 CTGGAGGCTGGGGCCAGAGAGTCTCTTTCACTGCAATGACAGAGCCCAAGGAGT 180
Db 392 CTGGAGGCTGGGGCCAGAGAGTCTCTTCCACATGATGTCACAGAGCCCTCAGTGGT 333

QY 181 GTTTCGCAAGGACAGTGTGGTGCATAATGAGTGGGACTGAGGAGGGCCAGTGGCCAC 240
Db 332 GTTTCGCAAGGACAGGACGACGACCAATGAGTGCAGCCATGGCCGAGCCAGTGGCCAC 273

QY 241 ACTGAGCTGTAGGTGGCCAGGCCCAGACAGAGGTGACGTGGTACAGGATGGGAAGAA 300
Db 272 AATGAGCTGTAGGTGGCCAGGCCCAGACAGAGGTGACATGTTACAGGATGGGAAGAA 213

QY 301 GCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGCTGGGGCTGCATGGCCAGCTGTGGT 360
Db 212 GCTGACTTCAAGCTCAAAAGTGGCTGTGGAGGCCCCCTGCAACCGAGGCTGTGGT 153

QY 361 GCAGAGCGAGGCGAGGACAGTGTGGGAGTACACTGTGAGGCTGGGGCCAGGCGGT 420
Db 152 GCAGAGCGGCGGCAAGGACAGATGATGGGAGTACAGCTGCGAGGCTGGGGGTGAGAAGGT 93

QY 421 CTCCTTCCACTGGATCTTTCAGAGCCCAAGGCGGTGTTTGCAGAGGAGCAGTGGCACA 480
Db 92 CTCCTTCCACTGGATCTTTCAGAGCCCTCAGTGGTGTGTTTGCAGAGGAGCAGCCATCAAG 33

QY 481 CAG 483
Db 32 CAG 30

RESULT 2
CF179218 621 bp mRNA linear EST 28-JUL-2003
LOCUS
DEFINITION
813316 MARC 3P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION
CF179218
VERSION
CF179218.1 GI:33250994
KEYWORDS
EST.
SOURCE
Sus scrofa (pig)
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 621)
Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,
Wise,T.A., Noneman,D.J., Wray,J.E. and Keefe,J.W.
A second set of porcine ESTs from a pooled-tissue normalized
library
Unpublished (2003)

CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross match v0.990329.
Plate: SRG8009 row: O column: 4
Seq primer: GTAATACGACTCAGCTATAGG.
Location/Qualifiers
1..621
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3P1G"
/note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

ORIGIN
Query Match 64.9%; Score 314.8; DB 14; Length 621;
Best Local Similarity 80.1%; Pred. No. 4.4e-59;
Matches 370; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1 GACGTGGTACAGGACGGGAGAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAA 60
Db 160 GACTTGGTACAGGATGGGAGAGAGCTGAGTTCAGCTCAAAAGTGGGTGGAGGCCAA 219

QY 61 GGGGTGCACACGAAGGCTGTGTGTCAGCAGAGGTGGGCAAGCAGATGCTGGGAGTACAG 120
Db 220 GGGGTGCACACCGAGGCTGTGTGTCATCAAGTGGGCAAGCAGATGCTGGGAGTACAG 279

QY 121 CTGGAGGCTGGGGCCAGAGAGTCTCTTTCACTGCAATGAGTGGGACTGAGGAGGGCCAGTGGCAGT 180
Db 280 CTGGAGTCTGGGGGTGATAAGGTCTCTTCCACTGGATGTCATATCCCTCAGTGGT 339

QY 181 GTTTCGCAAGGACAGTGTGGTGCATAATGAGTGGGACTGAGGAGGGCCAGTGGCCAC 240
Db 340 GTTTCGCAAGGACAGGACGACCATCAGCACTGAAGTCAAGCCATGGCCGAGCCAGTGGCCAC 399

QY 241 ACTGAGCTGTAGGTGGCCAGGCCCAGACAGAGGTGACGTGGTACAGGATGGGAAGAA 300
Db 400 ACTGAGCTGTAGGTGGGCCCATGCCAGACAGAGGTGACATGGTACAGGATGGGAAGAA 459

QY 301 GCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGCTGGGGCTGCATGGCGAGCTGTGGT 360
Db 460 GCTGACTTCAAGCTCAAAAGTGGCTGTGGAGGCCAATGGCTGCAACCGGATGCTGGTGGT 519

QY 361 GCAGAGCGAGGCGAGGACAGTGTGGGAGTACACTGTGAGGCTGGGGCCAGGCGGT 420
Db 520 GCATATGCGGGCAATGCATATGCTGGGAGTACAGCTGCGATGCTGGGGTCAATATGT 579

QY 421 CTCCTTCCACTGGATGTTTCAGAGCCCAAGGCGGTGTTTGC 462
Db 580 CTCCTTCCACTGGATGTTTCAGAGCCCTCAGTGGTGTGTTGC 621

RESULT 3
BG680164 1142 bp mRNA linear EST 01-MAY-2001
LOCUS
DEFINITION
602626828F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4751671 5',
mRNA sequence.
ACCESSION
BG680164
VERSION
BG680164.1 GI:13911561
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1. (bases 1 to 1142)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10608 row: h column: 08
High quality sequence stop: 708.
Location/Qualifiers
1. .1142
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:4751671"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP SKn4"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI CGAP Library."

```

ORIGIN
Query Match      44.5%; Score 216; DB 12; Length 1142;
Best Local Similarity 80.8%; Pred. No. 3.5e-37;
Matches 252; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 155 CTCACATCACAGAGCCCAAGCAGTGTTCCTCCNAGGACGAGTGGTGCATATAGAGTG 214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 322 CTCCTCTCCCAAGAGCCCAAGCGGTGTTTCTAAGGACGAGCCAGCACACAGGAGGTG 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 215 CGGACTGAGCAGGGGGCAGTGCCACACTGAGCTGTGAGTGGGCCAGGCCACACAGAG 274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 262 CAGCTGAACAGGGGCTACGGCCCACTGAGCTGTGAGTGGGCCAGGCCACACAGAG 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 275 GTCACTGTGTACAAGGATGGGAAGAAAGCTGAGTCTCAGTTTCGAAGATGCCATAGAGCT 334
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 202 GTCACTGTGTACAAGGACGGGAAGAGCTGAGCTCCAGCTTGAAGATGCATGTGAGGCT 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 335 GCGGGCTGCATCGGCAGCTGTGGTGCACAGGACGCCAGGCAGAGTCTGGGGAGTAC 394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 142 GCAGGCTGTACA CGGAGGCTGTGGTGCAGCAGGACGGCCAGGCAGACACCGGAGAGTAT 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 395 ACCTGTGAGGCTGGGGGCCAGCGGCTCTCCTTCCACCTGGATGTTTCAGAGCCCCAAGCGC 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 82 AGCTGTGAGGCTGGGGGCCAGCAGCTCTCCTTCGGCCTGCAGTGGCAGGTCAGTGCTT 23
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 455 GTGTTTGCAAAG 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22 GGGGATGCTGAG 11
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

| | | | | |
|------------|---|------------|--------|-----------------|
| RESULT 4 | AW898791 | 401 bp | linear | EST 24-MAY-2000 |
| LOCUS | AW898791 | | | |
| DEFINITION | CM0-NC0075-130400-332-f06 NC0075 Homo sapiens cDNA, mRNA sequence. | | | |
| ACCESSION | AW898791 | | | |
| VERSION | AW898791.1 | GI:8062996 | | |
| KEYWORDS | EST. | | | |
| SOURCE | Homo sapiens (human) | | | |
| ORGANISM | Homo sapiens | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | |
| REFERENCE | 1 (bases 1 to 401) | | | |
| AUTHORS | Dias Neto, E., Garcia Correa, R., Vertovski-Almeida, S., Briones, M.R., | | | |

Nagai, M.A., da Silva W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsumura, A., Bata, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20203663
PUBMED 10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil.
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL:
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1&t2=CMO-NN0075-130>)
400-332-f06&t3=2000-04-13&t4=1

| FEATURES | source | ORIGIN |
|--|--|----------------------------------|
| High quality sequence | 401. | |
| Location/Qualifiers | | |
| 1. .401 | | |
| /organism="Homo sapiens" | | |
| /mol_type="mRNA" | | |
| /db_xref="taxon:9606" | | |
| /dev_stage="Adult" | | |
| /clone_lib="NN0075" | | |
| /note="Organ: nervous normal; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 136,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." | | |
| Query Match | 43.9%; | Score 212.8; DB 10; Length 401; |
| Best Local Similarity | 79.1%; | Pred. No. 1.1e-36; |
| Matches 253; | Conservative 0; | Mismatches 67; Indels 0; Gaps 0; |
| 155 | CTGCACATCAGAGCCCAAGGACAGTGTTCGCAAGGACGATGTGTGTCATPAATGAGGTG | 214 |
| Db | | |
| 61 | CTGCTTTCCCGAGAGCCCAAGCGGTGTTCGCAAGGACGACGAGCAGGAGGAGGTG | 120 |
| Qy | | |
| 215 | CGGACTGAGGACGGGCGAGTGCACACTGAGCTGTGAGGTGGCCAGGCCCCAGACACAG | 274 |
| Db | | |
| 121 | CAGCTGAGCGGGACCATGTGCCAGCTGAGCTCGAGGTGGCCAGGCCCCAGACACAG | 180 |
| Qy | | |
| 275 | GTGACGTGTTACAAGATCGGAAGAGCTGAGCTCCAGTTCGAAAGTCGCATAGAGCT | 334 |
| Db | | |
| 181 | GTGACGTGTTACAAGAGCGGGAAGAACTGAGCTCCAGTTCGAAAGTCGGAATGAGGCC | 240 |
| Qy | | |
| 335 | GCGGGCTGCATGCGGCAGGCTGTGTGTGCAGCAGGCGAGGCCAGGACGATGCTGGGGAGTAC | 394 |
| Db | | |
| 241 | GTGGGCTGCACACGAGGCTGTGTGTGTGCAGGAGGCAGGCCAGGCGGAGCGGGAGTAC | 300 |
| Qy | | |
| 395 | ACCTGTGAGGCTGGGGGCCAGCGGCTCTCTTCCACCTGGATGTTTCAGAGCCCAAGCG | 454 |
| Db | | |
| 301 | AGCTGCAAGCCCGGGATCAGCGCTGTCCTTCCAGCTGCACGTGGCGAGTGCAGTGT | 360 |
| Qy | | |
| 455 | GTGTTTCAAAGGACGAGCT | 474 |
| Db | | |
| 361 | GGAGTCTCTGAGGAGCTCT | 380 |
| b | | |

[illegible]

Db 353 CCTCAGCTGTGAAGTAGCAGGATACACCGAAGTAAATGGTGTACAAAGATGAAGACT 412
 QY 301 GCTGAGCTCCAGTTTCGAAGTGCATAGAGGCTGCGGCTGCGATGCGGAGCTGGTGGT 360
 Db 413 GCTTGCCTCTCTAGAAAGTTCAAAATGGAACCTGTGGCAAAACCCGGCTCTGGTGT 472
 QY 361 GCAGAGCGAGCCAGCAGATGCTGGGAGTACACCTGTGAGCTGGGGCCAGCGGCT 420
 Db 473 GGAGCAGCTGGAGAAGAAGATGCTGGAGAATACCTGTGTGAGGCTCCCGCCAGAGGCT 532
 QY 421 CTCCTTCCACTGATTTTCAGAGCCCAAGGCGGTGTTTGCRAAGGA 468
 Db 533 GACCTTCAAGTTGGAAGCAACTGAACAGAGGCTAAATTTGAAAAGGA 580

RESULT 7
 AZ383233
 LOCUS 658 bp DNA linear GSS 02-OCT-2000
 DEFINITION IM0140117R Mouse 10kb plasmid UGCM library Mus musculus genomic
 clone UUGC1M0140117 R, genomic survey sequence.

ACCESSION AZ383233
 VERSION
 KEYWORDS GSS.

SOURCE
 ORGANISM Mus musculus (house mouse)

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (Bases 1 to 658)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
 Niederhausen, A., and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

JOURNAL
 COMMENT Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: dunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
 Plate: 0140 row: 1 column: 17
 Seq primer: CACACAGGAACACTATGACC
 Class: plasmid ends
 High quality sequence stop: 658.

FEATURES
 source
 1. .658
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0140117"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UGCM library"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (gi|4732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN

Query Match 40.2%; Score 194.8; DB 28; Length 658;
 Best Local Similarity 81.3%; Pred. No. 1.3e-32;
 Matches 226; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 QY 166 AGAGCCCAAGGCGAGTGTTCACAGGAGCAGTTGGTGCATATGAGTGGGACTGAGGC 225
 Db 182 AGACACAGGCTGATGTTTGCACAGAGCAGCAGCAGCAGGCTGAGGCTGAGGC 241
 QY 226 AGGGCCAGTGCACACTGAGCTGTGAGTGGCCAGCCAGAGGAGTACCTGTGTA 285
 Db 242 GGGGAACAGTGCACCTGAGCTGCGAGGTGGCCAGGCCAGACTGAGGTGACATGTT 301
 QY 286 CAGGATGGAGAGAGCTGAGCTCCAGTTCGAAAGTGGCATAGAGGCTGGGGTGCAT 345
 Db 302 CAAGGACGGGAAGAAGCTGAGCTCGAGCTCGAAGTGGCGTGGAGGCTTCGGGTGCTC 361
 QY 346 GCGGCAGCTGTGTGTGCAGCAGGCGAGCCAGCAGATGCTGGGAGTACACCTGTGAGGC 405
 Db 362 CAGGAGGCTGTGTGTGCAGCAGCGCGGCAAGSCGATGCTGGGAGTACAGCTGCGAGGC 421
 QY 406 TGGGGCCAGCGCGCTCTCTCCACTCGATGTTTCAG 443
 Db 422 CGGGGTCTAGAAAGCTCTCTCCGCTGGACGTGGCAG 459

RESULT 8

AI595154 574 bp mRNA linear EST 15-MAR-2000
 LOCUS ml39h10.y1 Stratagene mouse testis (#937308) Mus musculus cdna
 clone IMAGE:514435 5', mRNA sequence.

ACCESSION AI595154
 VERSION
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 574)
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
 Person, B., Swaller, I., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
 Ritter, B., Kohn, S., Shin, T., Jackson, J., Cardenas, M., McCann, R.,
 Waterston, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999

TITLE

Unpublished (1999)
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

JOURNAL

COMMENT Email: mouseest@wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)

MGI:308283
 Seq primer: -40RP from Gibco
 High quality sequence stop: 418
 POLYA=No.

FEATURES

source

1. .574
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CD-1"
 /db_xref="taxon:10090"
 /clone="IMAGE:514435"

PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@unist.ac.uk.

FEATURES

source
1. 588
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST98519"
/sex="Male and female"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEBEN11"
/note="Vector: pBluescript II KS(+); Site 1: EcoRI;
Site 2: NotI; This normalized library was constructed from
1 million independent clones. cDNA synthesis was initiated
using an oligo(dT) primer, using methylated C in the first
strand synthesis reaction. Following this first strand
reaction, double-stranded cDNA was blunted, ligated to
NotI adapters, digested with EcoRI, size-selected, and
cloned into the NotI and EcoRI compatible sites of a
custom modified MCS of the pBluescript (KS+) vector. The
library was normalized in 2 rounds using conditions
adapted from Soares et al., PNAS (1994) 91: 9228-9232 and
Bonaldo et al., Genome Research 6 (1996): 791, except that
a significantly longer reannealing hybridization was
used."

ORIGIN

Query Match 33.2%; Score 161.2; DB 13; Length 588;
Best Local Similarity 67.1%; Pred. No. 3.1e-25;
Matches 243; Conservative 0; Mismatches 118; Indels 1; Gaps 1;
Qy 77 CTGTTAGTGCAGCAGGTGGGCAAGCAGATGCTGGGAGTACACTCGAGGCTGGGGC 136
Db 1 CTGGTGGTGGAGCAGGTGGAGAAAGATGC-AGGGAGTACACTCGAGGCTGGTGGC 59
Qy 137 CAGAGAGTCTCTTCAACTGCACATCAGAGGCCCAAGGAGGTGTTGCCAAGGACAG 196
Db 60 CAGAACTGACCTTCAAGATTACTGTCTCAGAACGACGAGGTGTTTATCACAAGGAG 119
Qy 197 TTGTTGCAATATGAGTGGCAGTGAAGCAGGGGCCAGTGCCACACTGAGCTGTGAGGTG 256
Db 120 AAGGTGCAGAGAGGAGTGAAGGCTGCACCAACAGAAATGCCAGCTGAGTGCAGGTG 179
Qy 257 GCCCAGGCCACAGAGGTGACGTGTACAGGATGGGAGAGAGCTGAGCTCCAGTTGC 316
Db 180 GGCAGGAGAGACCGAGGTGAAGTGGTACAGGAGGGGAACTGATCCTCAGCAAG 239
Qy 317 AAAGTGGCATAGAGGCTGCGGGTGCATGCGGCAGCTGTGTTGTCAGCAGGAGGCCAG 376
Db 240 AAGTTGAGGTGGAGTCAAGGGGCAAACTGCTGGCTGGTGGTGGAGCCAGTGGAGA 299
Qy 377 GCAGATGCTGGGGAGTACACTGTGAGCTGGGGGCCAGCGGCTCTCTTCACTGAT 436
Db 300 AGAATGAGGGGAGTACACTGCGAGGCTGCTGGCCAGAAAGTGAAGCTTCAAGATTACT 359
Qy 437 GT 438
Db 360 GT 361

RESULT 11

BE159678/c
LOCUS BE159678 343 bp mRNA linear EST 21-JUN-2000
DEFINITION MRO-HT0407-180400-015-e07 HT0407 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE159678
VERSION BE159678.1 GI:8622399
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (Bases 1 to 343)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
MEDLINE
10737800
PUBMED
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project, this entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=MR0-HT0407-180
400-015-e07&t3=2000-04-18&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 343.

FEATURES

Location/Qualifiers
1. 343
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0407"
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 29.4%; Score 142.4; DB 10; Length 343;
Best Local Similarity 86.2%; Pred. No. 3.4e-21;
Matches 169; Conservative 0; Mismatches 26; Indels 1; Gaps 1;
Qy 163 CACAGAGCCCAAGGAGTGTGGCCAAAGGAGCAGTGTGTGCATAATGAGGTGCGACTGA 222
Db 228 CCAGAGCCCAAGTGTGTGTGGCCAAAGGAGCAGCAGCAGGAGGTGCAGGCTGA 169
Qy 223 GGCAGGGCCAGTGCACACTGAGCTGTGAGTGGCCAGGCCAGCCAGACAGAGTGCAGTG 282
Db 168 GCGGGGGCCAGTGCACCTGAGCTGCAGGTGGCCAGGCCAGGCCAGACAGAGTGCAGTG 109
Qy 283 GTACAGAGTGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCAGATAGAGGCTGCGGGCTG 342
Db 108 GTACAGAGTGGGAAGAGCTGAGTTCAGAGTCCAGAGTGGCGGTGA-CCCGTGGGCTG 50
Qy 343 CATGCGGCAGCTGGTG 358
Db 49 CACACGAGGCTGGTG 34

RESULT 12

AA061264/c
LOCUS AA061264 373 bp mRNA linear EST 03-FEB-1997
DEFINITION ml39h10.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
clone IMAGE:514435 5', similar to T7:G1017425 G1017425
IPISGKPLPKVTLSDRGVPLKATMEFNTETAEINTINKSVTADAGRYETTAANSSTGTTKAFINI
VVDRPPEPT

Wed Sep 22 12:33:12 2004

us-10-077-130-4_copy_3662_4146.rst

Page 10

Db 690 CCTGCGAAGCTGCTGGCCCAAAACTGACCTTCCAAATTACTGT 732

Search completed: September 19, 2004, 22:07:25
Job time : 1418.09 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:04:42 ; Search time 444 Seconds
(without alignments)
5070.582 Million cell updates/sec

Title: US-10-077-130-5

Perfect score: 41273

Sequence: 1 MDQPFSGAPRFLTRKAFV.....RNREKRALLYKSHNLQAVR 7968

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_29Jan04.*
- 1: Geneseq1980s.*
 - 2: Geneseq1990s.*
 - 3: Geneseq2000s.*
 - 4: Geneseq2001s.*
 - 5: Geneseq2002s.*
 - 6: Geneseq2003as.*
 - 7: Geneseq2003bs.*
 - 8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|---------|-------------|--------|-------|--------------------|
| 1 | 41273 | 100.0 | 7968 | 6 | ABG76187 Human ser |
| 2 | 13710 | 33.2 | 2630 | 6 | ABG76186 Human ser |
| 3 | 13528 | 32.8 | 2596 | 4 | AB30569 A splice |
| 4 | 10519.5 | 25.5 | 4691 | 5 | ABP70084 Human NOV |
| 5 | 10506 | 25.5 | 4675 | 5 | ABP70085 Human NOV |
| 6 | 9834.5 | 23.8 | 2328 | 6 | ABP58227 Human cel |
| 7 | 8423 | 20.4 | 1665 | 5 | AAO15372 Human myo |
| 8 | 8410 | 20.4 | 1665 | 5 | AAE24151 Human kin |
| 9 | 8407 | 20.4 | 1618 | 4 | AAE85504 Human pro |
| 10 | 8403 | 20.4 | 1610 | 4 | AB30568 A full le |
| 11 | 7073 | 17.1 | 1351 | 4 | AB30567 Amino aci |
| 12 | 4533 | 11.0 | 871 | 4 | AAE16274 Human kin |
| 13 | 4529 | 11.0 | 871 | 5 | AAE16274 Human kin |
| 14 | 3878.5 | 9.4 | 31267 | 6 | ABG74786 Human RGS |
| 15 | 3580.5 | 8.7 | 26926 | 4 | AAU05396 Human tit |
| 16 | 2447.5 | 5.9 | 548 | 4 | AB30571 A full le |
| 17 | 2415.5 | 5.9 | 548 | 4 | AB30572 A full le |
| 18 | 2360 | 5.7 | 4412 | 3 | AAV53666 Sequence |
| 19 | 2328.5 | 5.6 | 3267 | 7 | ADC99075 Human KPP |
| 20 | 2324 | 5.6 | 3208 | 7 | ABG74766 Human NOV |
| 21 | 2319 | 5.6 | 3268 | 7 | ABD79959 Human kin |
| 22 | 2270 | 5.5 | 3252 | 7 | ABG74764 Human NOV |
| 23 | 2203 | 5.3 | 3262 | 7 | ABD79961 Mouse ser |
| 24 | 2135 | 5.2 | 2380 | 5 | AAE19160 Human kin |
| 25 | 2127 | 5.2 | 2286 | 4 | AAE65635 Novel pro |

| | | | | | | | |
|----|--------|-----|------|---|----------|----------|-----------|
| 26 | 2102 | 5.1 | 5635 | 5 | ABP60991 | Abp60991 | Novel hum |
| 27 | 2085.5 | 5.1 | 3186 | 7 | AD47672 | AD47672 | Human NOV |
| 28 | 2067 | 5.0 | 2231 | 7 | AD879962 | AD879962 | Human KIA |
| 29 | 1840 | 4.5 | 3931 | 6 | ABU07377 | ABU07377 | Human pro |
| 30 | 1763.5 | 4.3 | 4796 | 4 | AB58665 | AB58665 | Drosophil |
| 31 | 1752 | 4.2 | 4495 | 6 | ABU69135 | ABU69135 | Human NOV |
| 32 | 1683 | 4.1 | 335 | 5 | ABP05662 | ABP05662 | Human ORF |
| 33 | 1553 | 3.8 | 1237 | 7 | ADD14123 | ADD14123 | Human src |
| 34 | 1487.5 | 3.6 | 7107 | 4 | AB58144 | AB58144 | Drosophil |
| 35 | 1249 | 3.0 | 3117 | 3 | AAV53667 | AAV53667 | Sequence |
| 36 | 1112 | 2.7 | 2173 | 5 | ABP69251 | ABP69251 | Human pol |
| 37 | 1081 | 2.6 | 4393 | 4 | AB331889 | AB331889 | Amino aci |
| 38 | 1081 | 2.6 | 4436 | 4 | ABG23265 | ABG23265 | Novel hum |
| 39 | 1074 | 2.6 | 4391 | 6 | AAE34390 | AAE34390 | Human per |
| 40 | 1072 | 2.6 | 209 | 3 | AAE40558 | AAE40558 | Human ORF |
| 41 | 1072 | 2.6 | 209 | 5 | ABP31663 | ABP31663 | Human ORF |
| 42 | 1025 | 2.5 | 1953 | 5 | AAU84351 | AAU84351 | Protein M |
| 43 | 1018 | 2.5 | 5701 | 4 | AB36684 | AB36684 | Peptide # |
| 44 | 1018 | 2.5 | 5701 | 4 | AB222021 | AB222021 | Protein # |
| 45 | 1018 | 2.5 | 5701 | 4 | AAE69843 | AAE69843 | Human bon |

ALIGNMENTS

RESULT 1

ABG76187 standard; protein; 7968 AA.

XX
AC ABG76187;
XX
XX
DT 09-MAY-2003 (first entry)
XX
XX
DE Human serine/threonine or protein kinase 12599.

XX Human; enzyme; serine/threonine kinase; protein kinase; 12599;
KW cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma; immunogen;
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
KW haemolytic anaemia; cellular proliferative disorder; cancer;
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.

XX Homo sapiens.

XX US2002168742-A1.

XX 14-NOV-2002.

XX 15-FEB-2002; 2002US-00077130.

XX 15-FEB-2001; 2001US-0269201P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Kapeller-Libermann R, Acton SL;

XX WPI; 2003-298729/29.

XX N-PSDB; ABX11642.

XX Novel isolated human protein kinase, designated 59079 or 12599 polypeptide, useful as diagnostic and therapeutic agents for preventing cardiovascular diseases, proliferative disorders, and protein kinase disorders.

XX Claim 8; Page 84-104; 119pp; English.

XX The invention relates to an isolated human serine/threonine or protein kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule comprising at least 85% identity to the nucleic acids appearing as CC AEX11641 and AEX11642 or their complement, a naturally occurring variant CC of the kinases or their fragments. Also included are a non-human host

cell containing the nucleic acids, an antibody specific for the proteins, identifying a compound which binds to the kinase (by contacting the kinase or a cell expressing the kinase with a test compound and determining whether the kinase binds to the test compound) and modulating the activity of kinase using the identified compound. The kinases and their encoding nucleic acids are useful as diagnostic and therapeutic agents for preventing a disease or condition associated with an aberrant or unwanted 59079 or 12599 activity in a subject, including cardiovascular diseases such as heart failure, and myocardial infarction; disorders involving blood vessels such as atherosclerosis, and Kaposi's sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia, Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders such as cancer; and protein kinase disorders such as autoimmune disorders, diabetes mellitus, psoriasis, inflammatory bowel disease, rheumatoid arthritis, and multiple sclerosis (many examples of diseases and disorders are included in the specification). The kinases, their encoding nucleic acids and antibodies are useful in screening assays, detection assays (e.g. forensic biology), and predictive medicine (e.g. diagnostic assays, prognostic assays, and monitoring clinical trials and pharmacogenomics). The kinases and their encoding nucleic acids are useful as query sequences to perform a search against public databases to identify other family members or related sequences. The present sequence represents kinase 12599

Sequence 7968 AA;

Query Match 100.0%; Score 41273; DB 6; Length 7968;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 7968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDQPFSGAPFLTRPFAFVSVGKDATLSQIVGNPTPOVSWEKDOQPTAGARFLAQ 60
 DB 1 MDQPFSGAPFLTRPFAFVSVGKDATLSQIVGNPTPOVSWEKDOQPTAGARFLAQ 60
 QY 61 DGDLYRLTILDALGDSQYVCRARNAIGFAFAVGLQVDAEACAEQAFHLLRPTSIR 120
 DB 61 DGDLYRLTILDALGDSQYVCRARNAIGFAFAVGLQVDAEACAEQAFHLLRPTSIR 120
 QY 121 VREGSEATFCRVGSGPRPAVSWGKGRRLGEPDGPVRVVEELGEASALRIARPRDGG 180
 DB 121 VREGSEATFCRVGSGPRPAVSWGKGRRLGEPDGPVRVVEELGEASALRIARPRDGG 180
 QY 181 TYEVAENPLGAASAAALVVDSDAATASRPGTSTAALLAHQLQRREARAEAPASPP 240
 DB 181 TYEVAENPLGAASAAALVVDSDAATASRPGTSTAALLAHQLQRREARAEAPASPP 240
 QY 241 STGTRCTVTGKHARLSCYVTPGPKPETVWKDGLVTEGRRHVYVEDAENFVLKILF 300
 DB 241 STGTRCTVTGKHARLSCYVTPGPKPETVWKDGLVTEGRRHVYVEDAENFVLKILF 300
 QY 301 CKQSDRGLYCTASNLVQGTYSVVLVVRPAVPFVKRLQDLEVRKESATFLCEVPQPS 360
 DB 301 CKQSDRGLYCTASNLVQGTYSVVLVVRPAVPFVKRLQDLEVRKESATFLCEVPQPS 360
 QY 361 TEAAWFKEETRLMASAKYIEEBGTERRLTVRNVSADDDAVYICETPEGSTVAELAVQG 420
 DB 361 TEAAWFKEETRLMASAKYIEEBGTERRLTVRNVSADDDAVYICETPEGSTVAELAVQG 420
 QY 421 NLLRKLPRKTRAVRGDTAMFCVELAVPVPVHMLRNOEEVVAGRVASAEGRTHLTIS 480
 DB 421 NLLRKLPRKTRAVRGDTAMFCVELAVPVPVHMLRNOEEVVAGRVASAEGRTHLTIS 480
 QY 481 QCCLEDVGQAFMAGDCQSTFRFCVSAPRKPPLPPVDPVVKARMSSSVILSNPPHGE 540
 DB 481 QCCLEDVGQAFMAGDCQSTFRFCVSAPRKPPLPPVDPVVKARMSSSVILSNPPHGE 540
 QY 541 RPVTIDGVLVEKKLGYTWLRCHAEAWATPELTVDADVAEENFQFRVSALNSFGOSPY 600
 DB 541 RPVTIDGVLVEKKLGYTWLRCHAEAWATPELTVDADVAEENFQFRVSALNSFGOSPY 600
 QY 601 LEFPGTVHLAPKLAVRTPLKAVQAVEGGEVTFSDLTVASAGBWFLDGQALKASSVYEIH 660
 DB 601 LEFPGTVHLAPKLAVRTPLKAVQAVEGGEVTFSDLTVASAGBWFLDGQALKASSVYEIH 660

QY 661 CDTRHTLTIREVPASLHGAQLKFVANGIESSIRMEVRAAPGLTANKFPAAAAAEVLARL 720
 DB 661 CDTRHTLTIREVPASLHGAQLKFVANGIESSIRMEVRAAPGLTANKFPAAAAAEVLARL 720
 QY 721 HEEAQLLELSQAAAATWLVKDGRTLSFGPKYEVQASAGRRVLLVRDVARDADAGLYECVS 780
 DB 721 HEEAQLLELSQAAAATWLVKDGRTLSFGPKYEVQASAGRRVLLVRDVARDADAGLYECVS 780
 QY 781 RGRITAYQLSVQGLARFLHKDMAGSCVDVAGGPAQFECETSEAHVHVHWYKDMELGHS 840
 DB 781 RGRITAYQLSVQGLARFLHKDMAGSCVDVAGGPAQFECETSEAHVHVHWYKDMELGHS 840
 QY 841 GERFLQEDVGTTRHRLVAATVTRQDEGTYSRCVGSVDVDFRLRVSEPKVVFVFAKEQLARRKL 900
 DB 841 GERFLQEDVGTTRHRLVAATVTRQDEGTYSRCVGSVDVDFRLRVSEPKVVFVFAKEQLARRKL 900
 QY 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVKCMEATGCTTRRLVVQAGADAGEY 960
 DB 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVKCMEATGCTTRRLVVQAGADAGEY 960
 QY 961 SCEAGGQRLSFLHDYKEPKVVFVFAKQDVAAHSEVQAEAGANATLSCEVAQAQAEVVMYKDGK 1020
 DB 961 SCEAGGQRLSFLHDYKEPKVVFVFAKQDVAAHSEVQAEAGANATLSCEVAQAQAEVVMYKDGK 1020
 QY 1021 KLSSSLKVHVEAKGCRRLVVQAGKTDAGDYSCEARGQVSRFLHITTEPKMFAKQSV 1080
 DB 1021 KLSSSLKVHVEAKGCRRLVVQAGKTDAGDYSCEARGQVSRFLHITTEPKMFAKQSV 1080
 QY 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSKVKCMEATGCTTRRLVVQAGKAD 1140
 DB 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSKVKCMEATGCTTRRLVVQAGKAD 1140
 QY 1141 AGEYSCEAGGQVSRVPHLHI TEFGVFAKEQSVHNEVQAEAGTGTAMLSCEVAQAQTEVTWY 1200
 DB 1141 AGEYSCEAGGQVSRVPHLHI TEFGVFAKEQSVHNEVQAEAGTGTAMLSCEVAQAQTEVTWY 1200
 QY 1201 KDGKLSKSSKVRMEVKGCTTRRLVVQVQKADAGEYSCEAGGQVSRVFLHITTEPKVFAK 1260
 DB 1201 KDGKLSKSSKVRMEVKGCTTRRLVVQVQKADAGEYSCEAGGQVSRVFLHITTEPKVFAK 1260
 QY 1261 EQLVHNEVTRTAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVRIEAAGCMRLVQQA 1320
 DB 1261 EQLVHNEVTRTAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVRIEAAGCMRLVQQA 1320
 QY 1321 GQADAGEYTCAGGQRLSFHLDVSEPKVFAKEQLAHRKVAEAGATATLSCEVAQAQTE 1380
 DB 1321 GQADAGEYTCAGGQRLSFHLDVSEPKVFAKEQLAHRKVAEAGATATLSCEVAQAQTE 1380
 QY 1381 VTWYKDGKLSKSSKVRMEVKGCTTRRLVVQVQACQADTGEYSCEAGGQRLSFLDVASPKV 1440
 DB 1381 VTWYKDGKLSKSSKVRMEVKGCTTRRLVVQVQACQADTGEYSCEAGGQRLSFLDVASPKV 1440
 QY 1441 VFAKEQPHREVVQAGAGASTTILSCEVAQAQTEVVMYKDGKLSKSSKVRMEVAGCTRLV 1500
 DB 1441 VFAKEQPHREVVQAGAGASTTILSCEVAQAQTEVVMYKDGKLSKSSKVRMEVAGCTRLV 1500
 QY 1501 VQAGQADAGEYSCEAGSQRILSFHLHVAEPKAVFAKEQSPASREVQAEAGTATLSCEVAQ 1560
 DB 1501 VQAGQADAGEYSCEAGSQRILSFHLHVAEPKAVFAKEQSPASREVQAEAGTATLSCEVAQ 1560
 QY 1561 AQTEVTWYKDGKLSKSSKVRMEVAGCTTRRLVVQAGADAGEYSCAGQDRLSFHLHVA 1620
 DB 1561 AQTEVTWYKDGKLSKSSKVRMEVAGCTTRRLVVQAGADAGEYSCAGQDRLSFHLHVA 1620
 QY 1621 BPKVFAKEQPAHREVVQAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVRVEAVGCT 1680
 DB 1621 BPKVFAKEQPAHREVVQAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVRVEAVGCT 1680
 QY 1681 RRLVVQAGQADAGEYSCEAGGQRLSFLHVAELBPOISERPCCRRLVVKHEHDIILTA 1740
 DB 1681 RRLVVQAGQADAGEYSCEAGGQRLSFLHVAELBPOISERPCCRRLVVKHEHDIILTA 1740

QY 1741 TLATPSAATVTLKDGVEIRSKHETASQDTHTLTVHGAOVLDSAIYSCRVAEGQDF 1800
Db 1741 TLATPSAATVTLKDGVEIRSKHETASQDTHTLTVHGAOVLDSAIYSCRVAEGQDF 1800
QY 1801 PVQVEEAAKFCRLLEPVCGELGTVTLACELSPACAEVVMRCNTQPRVKGKPFQWVAEG 1860
Db 1801 PVQVEEAAKFCRLLEPVCGELGTVTLACELSPACAEVVMRCNTQPRVKGKPFQWVAEG 1860
QY 1861 PVRSLTVLGLRAEDAGEVVCESRDHHTSAOLTVSPVVKFMSGLSTVVAEEGGEATFQC 1920
Db 1861 PVRSLTVLGLRAEDAGEVVCESRDHHTSAOLTVSPVVKFMSGLSTVVAEEGGEATFQC 1920
QY 1921 VVSPSDVAVVMFRDGLLOPSEKFAISQSGASHSLTISDLVLEDAGQITVVEAGASSAA 1980
Db 1921 VVSPSDVAVVMFRDGLLOPSEKFAISQSGASHSLTISDLVLEDAGQITVVEAGASSAA 1980
QY 1981 LRVREAPVLFKKLEPQTVBERSSVTLEVLTRPKBELRWTRNATALAPGNVEIHAEGA 2040
Db 1981 LRVREAPVLFKKLEPQTVBERSSVTLEVLTRPKBELRWTRNATALAPGNVEIHAEGA 2040
QY 2041 RHRLVLHNVGFADRGFFCGCEPDDKTQAKLTVENRQVRLVRGLQAVEAREQGTTATMEVQL 2100
Db 2041 RHRLVLHNVGFADRGFFCGCEPDDKTQAKLTVENRQVRLVRGLQAVEAREQGTTATMEVQL 2100
QY 2101 SHADVGSWTRDGLRFQOGPTCHLAVRGPMTLTLISGLRPEDSGIMVFKAEVHTSARLV 2160
Db 2101 SHADVGSWTRDGLRFQOGPTCHLAVRGPMTLTLISGLRPEDSGIMVFKAEVHTSARLV 2160
QY 2161 VTELVPFSRPLQDVVTTKEKVTLECELSRPNDVRWLKDGVELRAGKTMALAAQACR 2220
Db 2161 VTELVPFSRPLQDVVTTKEKVTLECELSRPNDVRWLKDGVELRAGKTMALAAQACR 2220
QY 2221 SLTIYRCEFFADQGVVCDHAQAQSSASVKVQGRVTYTLIYRRVLAEDAGEIQFVAENAESR 2280
Db 2221 SLTIYRCEFFADQGVVCDHAQAQSSASVKVQGRVTYTLIYRRVLAEDAGEIQFVAENAESR 2280
QY 2281 ACLRVKELPVTILVRPLRDKIAMEKHGVLCEQVSPASQVRFKGSQELQPGPKYELVSD 2340
Db 2281 ACLRVKELPVTILVRPLRDKIAMEKHGVLCEQVSPASQVRFKGSQELQPGPKYELVSD 2340
QY 2341 GLYRKLIISDVHAEDEDYTCADGVKTSQAQFFVEEQSIITIVRGLQDVTVMEPAPAMFEC 2400
Db 2341 GLYRKLIISDVHAEDEDYTCADGVKTSQAQFFVEEQSIITIVRGLQDVTVMEPAPAMFEC 2400
QY 2401 ETSIPSVRPPKWLIGKTVLQAGNVGLBOGTVHRLMLRRTCTMTGPVHFTVCKSRSSA 2460
Db 2401 ETSIPSVRPPKWLIGKTVLQAGNVGLBOGTVHRLMLRRTCTMTGPVHFTVCKSRSSA 2460
QY 2461 RLWVSDIIPVLTPLPKTGRELOSIVLSCDFRPAPKAVQYKDDTPLSPSEKPKMSLEG 2520
Db 2461 RLWVSDIIPVLTPLPKTGRELOSIVLSCDFRPAPKAVQYKDDTPLSPSEKPKMSLEG 2520
QY 2521 QMAELRIILRLMPADAGVYRCQAGSAHSSTEVTVBEAREVTVTGPIQDAEATEEGWASFSCE 2580
Db 2521 QMAELRIILRLMPADAGVYRCQAGSAHSSTEVTVBEAREVTVTGPIQDAEATEEGWASFSCE 2580
QY 2581 LSHDEEVEWSLNGPLVNDSFHETSHKGRHRTLVLKSIQADAGIVRASLSKYVSTSARL 2640
Db 2581 LSHDEEVEWSLNGPLVNDSFHETSHKGRHRTLVLKSIQADAGIVRASLSKYVSTSARL 2640
QY 2641 EVRVKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRKDXGVLQGPSDKYDFLHTAGT 2700
Db 2641 EVRVKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRKDXGVLQGPSDKYDFLHTAGT 2700
QY 2701 RGLVVDHVSPEAGLYTCHVGSEETRARVRVHDHLVHGITKLTMEVLEGESCSFEVCVLS 2760
Db 2701 RGLVVDHVSPEAGLYTCHVGSEETRARVRVHDHLVHGITKLTMEVLEGESCSFEVCVLS 2760
QY 2761 HESASDPAMWTVGGKTVGSSSRFQATROGRKYILVREAAPSDAGEVVVFSVRGLITSKASL 2820
Db 2761 HESASDPAMWTVGGKTVGSSSRFQATROGRKYILVREAAPSDAGEVVVFSVRGLITSKASL 2820
QY 2821 IVRERPAALIKPLEDQWVAPGEDVELRCELSSRAGTFVHMLKDKRAIKRSQKYDVVCEGTM 2880

Db 2821 IVRERPAALIKPLEDQWVAPGEDVELRCELSSRAGTFVHMLKDKRAIKRSQKYDVVCEGTM 2880
QY 2881 AMLVTRGASLKDAGEYTCBEVASKSTASLHVEKANCFTTEELTNLQVEEKGTAFTCKTKE 2940
Db 2881 AMLVTRGASLKDAGEYTCBEVASKSTASLHVEKANCFTTEELTNLQVEEKGTAFTCKTKE 2940
QY 2941 HPAATVTVWKGLLELRASGHKQFQSEGLTLRLTITISALEKADSDTYTCDIGQASRAQLIV 3000
Db 2941 HPAATVTVWKGLLELRASGHKQFQSEGLTLRLTITISALEKADSDTYTCDIGQASRAQLIV 3000
QY 3001 QGRRVHIIBDLBDVDVQEGSSATFRCRISPANYPVPHFPLDKTPHLANELNBIIDAPQGY 3060
Db 3001 QGRRVHIIBDLBDVDVQEGSSATFRCRISPANYPVPHFPLDKTPHLANELNBIIDAPQGY 3060
QY 3061 HVLTLRQLALKDSGTIYFAGDORASAAALRVTEKSPVSRELTDAITIEGEBDLTVCEFS 3120
Db 3061 HVLTLRQLALKDSGTIYFAGDORASAAALRVTEKSPVSRELTDAITIEGEBDLTVCEFS 3120
QY 3121 TCDIPMCWTGDKGTLRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRV 3180
Db 3121 TCDIPMCWTGDKGTLRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRV 3180
QY 3181 HARPVRFQALKDLEVLGGAATLRCVLSSVAAPVKWCYGNVLPDGYSLRQEGAMLE 3240
Db 3181 HARPVRFQALKDLEVLGGAATLRCVLSSVAAPVKWCYGNVLPDGYSLRQEGAMLE 3240
QY 3241 LVVNRNLRPDGSGRYSCSGFDQTTSATLTVTALPAQFIGLKRNEKATEGATATLRCELSKT 3300
Db 3241 LVVNRNLRPDGSGRYSCSGFDQTTSATLTVTALPAQFIGLKRNEKATEGATATLRCELSKT 3300
QY 3301 APVWRKGSSETLRDGDYCLRDQGANCELOIIRGLAMVDAAEYSCVCGEERTSASLTIRM 3360
Db 3301 APVWRKGSSETLRDGDYCLRDQGANCELOIIRGLAMVDAAEYSCVCGEERTSASLTIRM 3360
QY 3361 PAHFIOGRHQSIEGATATLRCELSKAAPVWRKGRSLRDGRHSIRDQAVCELOIC 3420
Db 3361 PAHFIOGRHQSIEGATATLRCELSKAAPVWRKGRSLRDGRHSIRDQAVCELOIC 3420
QY 3421 GLAVADAGYSVCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWCELSKVAPVE 3480
Db 3421 GLAVADAGYSVCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWCELSKVAPVE 3480
QY 3481 WRKGPENLRDGRYTLRQSGTRCELOICGLAMADAGEYLCVCGOERTSATLIRALPARF 3540
Db 3481 WRKGPENLRDGRYTLRQSGTRCELOICGLAMADAGEYLCVCGOERTSATLIRALPARF 3540
QY 3541 IEDVKNOEAREGATAVLOCELNSAAPVWRKGSSETLRDGDYSLRDQGTKCELOIRGLAM 3600
Db 3541 IEDVKNOEAREGATAVLOCELNSAAPVWRKGSSETLRDGDYSLRDQGTKCELOIRGLAM 3600
QY 3601 ADTGEYSVCVCGOERTSATLTVRALPIKFTTEGLRNEEATEGATAYLRCELSKMAPVEMWK 3660
Db 3601 ADTGEYSVCVCGOERTSATLTVRALPIKFTTEGLRNEEATEGATAYLRCELSKMAPVEMWK 3660
QY 3661 HETLRDGRHSIRDQGRARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSPFIEGL 3720
Db 3661 HETLRDGRHSIRDQGRARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSPFIEGL 3720
QY 3721 RNEEATEGDTATLWCELSKAAAPVWRKKGHETIRDGRHSIRDQGRCELOIRGLAVVDAG 3780
Db 3721 RNEEATEGDTATLWCELSKAAAPVWRKKGHETIRDGRHSIRDQGRCELOIRGLAVVDAG 3780
QY 3781 EYSVCVCGOERTSATLTVRALPARFIEDVKNOEAREGATAVLOCELNSKAAPVWRKGSSETL 3840
Db 3781 EYSVCVCGOERTSATLTVRALPARFIEDVKNOEAREGATAVLOCELNSKAAPVWRKGSSETL 3840
QY 3841 RGGDRYSLRDQGTGRCELOIRHGLSVADTGEYSVCVCGOERTSATLTVRAPQVFPREPLQSIQ 3900
Db 3841 RGGDRYSLRDQGTGRCELOIRHGLSVADTGEYSVCVCGOERTSATLTVRAPQVFPREPLQSIQ 3900
QY 3901 AEEGSGTATLOCELSEBTATVWVSKGLOIQAANGRAEPLOGCTAEALVLODLOREDGEXT 3960

| | | | |
|----|------|--|------|
| Db | 3901 | ABEGSTATLQCELSSEPTATVVMWSKGLQLQANGRRPRLOCCTAELVLQDLQREDTGEYT | 3960 |
| Qy | 3961 | CTCSQATSATLVTVAAPVRELRLELQHQVEDEGCTAHLCCELSRAGASVEMRKGSLQFF | 4020 |
| Db | 3961 | CTCSQATSATLVTVAAPVRELRLELQHQVEDEGCTAHLCCELSRAGASVEMRKGSLQFF | 4020 |
| Qy | 4021 | CAKTMQVODGAAAEELLVRGVQEBDAGDYTCDTGHTQSMASLSVVRPRPKTRILQSLQE | 4080 |
| Db | 4021 | CAKTMQVODGAAAEELLVRGVQEBDAGDYTCDTGHTQSMASLSVVRPRPKTRILQSLQE | 4080 |
| Qy | 4081 | TGDIARLCCQLSDAESGAVVQWLKXGVHLAGPKYEMRSOCATRELLIHOLEAKDTGEYA | 4140 |
| Db | 4081 | TGDIARLCCQLSDAESGAVVQWLKXGVHLAGPKYEMRSOCATRELLIHOLEAKDTGEYA | 4140 |
| Qy | 4141 | CVTGQKTAASLRVTEPEVTIVRGVDAEVTADDEVEFCEVSRAGATGVQWCLQGLPLQ | 4200 |
| Db | 4141 | CVTGQKTAASLRVTEPEVTIVRGVDAEVTADDEVEFCEVSRAGATGVQWCLQGLPLQ | 4200 |
| Qy | 4201 | SNEVTEAVRDRGRLHTLRKGVTPEDAGTYSFHLGNHASSAQLTVRAPEVTILEPLQDVQ | 4260 |
| Db | 4201 | SNEVTEAVRDRGRLHTLRKGVTPEDAGTYSFHLGNHASSAQLTVRAPEVTILEPLQDVQ | 4260 |
| Qy | 4261 | LSEGQDASFOCLRSLRASQEARWALGGVPLQANEMNDITVEQGTJLHLLTLHKVTLDEAGT | 4320 |
| Db | 4261 | LSEGQDASFOCLRSLRASQEARWALGGVPLQANEMNDITVEQGTJLHLLTLHKVTLDEAGT | 4320 |
| Qy | 4321 | VSFHVGTCSSEAOIKVTAKNVTVRGLENVBALEGEALFECOLSOPEVAATHLLDDEPV | 4380 |
| Db | 4321 | VSFHVGTCSSEAOIKVTAKNVTVRGLENVBALEGEALFECOLSOPEVAATHLLDDEPV | 4380 |
| Qy | 4381 | RTSENAEVVFPENGRLHLLKNLRPODSRVTPFAGDMVTSAPFLTVRGWRLEILPLKN | 4440 |
| Db | 4381 | RTSENAEVVFPENGRLHLLKNLRPODSRVTPFAGDMVTSAPFLTVRGWRLEILPLKN | 4440 |
| Qy | 4441 | AAVRAGAQAARTCTLSEAVPVGEASWYINGAAVOPDSDSMTVTADGSHQALLRSQPHH | 4500 |
| Db | 4441 | AAVRAGAQAARTCTLSEAVPVGEASWYINGAAVOPDSDSMTVTADGSHQALLRSQPHH | 4500 |
| Qy | 4501 | AGEVTFACRDVAVASARLTVGLPDPPEDAEVVAHSHHTVLSWAAPMSDGGGLCGYRVE | 4560 |
| Db | 4501 | AGEVTFACRDVAVASARLTVGLPDPPEDAEVVAHSHHTVLSWAAPMSDGGGLCGYRVE | 4560 |
| Qy | 4561 | VKEGATGWRILCHELVPGEPCVWDLAPGETYFRVAAVGPVAGPEVHLPTQVRLAEP | 4620 |
| Db | 4561 | VKEGATGWRILCHELVPGEPCVWDLAPGETYFRVAAVGPVAGPEVHLPTQVRLAEP | 4620 |
| Qy | 4621 | KPVPPQSPAPESROVAAAGEDVSLEVVAAAGEVIMHKGMERIQPGRFVVSQGRQOVL | 4680 |
| Db | 4621 | KPVPPQSPAPESROVAAAGEDVSLEVVAAAGEVIMHKGMERIQPGRFVVSQGRQOVL | 4680 |
| Qy | 4681 | VIKGFTAEQDGEYHCGLAQSGSICPAATFOVALSPASVDEAPQSLPPEAAQSGDLHLW | 4740 |
| Db | 4681 | VIKGFTAEQDGEYHCGLAQSGSICPAATFOVALSPASVDEAPQSLPPEAAQSGDLHLW | 4740 |
| Qy | 4741 | BALARKRMSREPTLDSISELPEEDGSRQRLPQAEAEVADPLSEGYSSTADELARTGDAL | 4800 |
| Db | 4741 | BALARKRMSREPTLDSISELPEEDGSRQRLPQAEAEVADPLSEGYSSTADELARTGDAL | 4800 |
| Qy | 4801 | SHTSSDDESAGTSLVTYLLKAGRPCTSPASKVGAAPAPSVPKPOQQQEPPLAARPPPLG | 4860 |
| Db | 4801 | SHTSSDDESAGTSLVTYLLKAGRPCTSPASKVGAAPAPSVPKPOQQQEPPLAARPPPLG | 4860 |
| Qy | 4861 | DLSTKDLGDPMSDAAVKIQAAFKGYKVRKEMKQOEGPMFSHTFGDTEAQVGDALRECV | 4920 |
| Db | 4861 | DLSTKDLGDPMSDAAVKIQAAFKGYKVRKEMKQOEGPMFSHTFGDTEAQVGDALRECV | 4920 |
| Qy | 4921 | VASKADYRARWLKGVVELTDCGRHHHIDQLGDGTCSLLIAGLDRADAGCYTCQVSNKFGOV | 4980 |
| Db | 4921 | VASKADYRARWLKGVVELTDCGRHHHIDQLGDGTCSLLIAGLDRADAGCYTCQVSNKFGOV | 4980 |
| Qy | 4981 | THSACVVVSGSESAESSSGGELDDAFRRARLRLHRLFRTKSPAEVSDDEELFUSADEGPA | 5040 |
| Db | 4981 | THSACVVVSGSESAESSSGGELDDAFRRARLRLHRLFRTKSPAEVSDDEELFUSADEGPA | 5040 |
| Qy | 5041 | EPEEPADQTYREDDEHFCIRFEALTEARQAVTRFQEMFATLIGIVEIKLVEQGPRRVEM | 5100 |
| Db | 5041 | EPEEPADQTYREDDEHFCIRFEALTEARQAVTRFQEMFATLIGIVEIKLVEQGPRRVEM | 5100 |
| Qy | 5101 | CISKSTPAPVTPPEPLPSILLTSDAAPVFLTELQNOVEODGVYVSDCVVTCQMPSVRWPF | 5160 |
| Db | 5101 | CISKSTPAPVTPPEPLPSILLTSDAAPVFLTELQNOVEODGVYVSDCVVTCQMPSVRWPF | 5160 |
| Qy | 5161 | KDGKLEEDDDHYMINEDQGGHQLIITAVVPADGVYRCLAEENSGVSSSTKAEILRVDLTS | 5220 |
| Db | 5161 | KDGKLEEDDDHYMINEDQGGHQLIITAVVPADGVYRCLAEENSGVSSSTKAEILRVDLTS | 5220 |
| Qy | 5221 | TDYDTAADAATSSSYFSAQGYLSREQGTSTTDEGQLPOVVEELRDLQVAPGTRLAKF | 5280 |
| Db | 5221 | TDYDTAADAATSSSYFSAQGYLSREQGTSTTDEGQLPOVVEELRDLQVAPGTRLAKF | 5280 |
| Qy | 5281 | OLKVKGYPAPRLYWFKDGQPLTASAHIRMTCKILHTLEIISVTREDSGOVAAVINSAMG | 5340 |
| Db | 5281 | OLKVKGYPAPRLYWFKDGQPLTASAHIRMTCKILHTLEIISVTREDSGOVAAVINSAMG | 5340 |
| Qy | 5341 | AAYSARLLVRGPDPEPEKPSADVHEQLVPPRMLERFTPKVKKGSSITFVSVKVEGRPVP | 5400 |
| Db | 5341 | AAYSARLLVRGPDPEPEKPSADVHEQLVPPRMLERFTPKVKKGSSITFVSVKVEGRPVP | 5400 |
| Qy | 5401 | TVHMLREAEARGVLMWIGPDTFGYTVASSAQOHSVLVDVGRQHGTYCTIASNAAGQALC | 5460 |
| Db | 5401 | TVHMLREAEARGVLMWIGPDTFGYTVASSAQOHSVLVDVGRQHGTYCTIASNAAGQALC | 5460 |
| Qy | 5461 | SASLHVSGLPKVEGEKVKKEALISTFLOQTTOAISAOGLETPASADLGGQKESPLAAKE | 5520 |
| Db | 5461 | SASLHVSGLPKVEGEKVKKEALISTFLOQTTOAISAOGLETPASADLGGQKESPLAAKE | 5520 |
| Qy | 5521 | ALGHLSLAEVCTEPELOKLTSOITEMVSAKLTQAKLOVPGGSDSDSKTSPASPRHGRSR | 5580 |
| Db | 5521 | ALGHLSLAEVCTEPELOKLTSOITEMVSAKLTQAKLOVPGGSDSDSKTSPASPRHGRSR | 5580 |
| Qy | 5581 | PSSSTIQESSSESSEDGDARGEIFDIYVVTADYPLGAEQDAITLREGQVVEVLDAAHPLRW | 5640 |
| Db | 5581 | PSSSTIQESSSESSEDGDARGEIFDIYVVTADYPLGAEQDAITLREGQVVEVLDAAHPLRW | 5640 |
| Qy | 5641 | LVRTPTKSSPSRQGWVSFAYLDRRLKLSPEWGAABAPPEPCEAVSDEDEYKARLSVIOE | 5700 |
| Db | 5641 | LVRTPTKSSPSRQGWVSFAYLDRRLKLSPEWGAABAPPEPCEAVSDEDEYKARLSVIOE | 5700 |
| Qy | 5701 | LLSSBOAFVBELOFLQSHHLOHLERCPRVPIAVAGQKAVIPRNVRDITGRHSSFLQELQ | 5760 |
| Db | 5701 | LLSSBOAFVBELOFLQSHHLOHLERCPRVPIAVAGQKAVIPRNVRDITGRHSSFLQELQ | 5760 |
| Qy | 5761 | CTDDDDVAMCTIKQAAFEQYLEFLVGRVQAESVVVSTAIQEFYKKYAEAEALLAGDSPSQ | 5820 |
| Db | 5761 | CTDDDDVAMCTIKQAAFEQYLEFLVGRVQAESVVVSTAIQEFYKKYAEAEALLAGDSPSQ | 5820 |
| Qy | 5821 | PPPLQHYLEOPVERVQRYQALLKELIRNKARNONCALLBOAVVVSALPQRAENKLHV | 5880 |
| Db | 5821 | PPPLQHYLEOPVERVQRYQALLKELIRNKARNONCALLBOAVVVSALPQRAENKLHV | 5880 |
| Qy | 5881 | SIMENYPTLEALGEPTRQGHFI VMEGAPGARMKPNHNRHVFLFRNHLVTCRPRDSRT | 5940 |
| Db | 5881 | SIMENYPTLEALGEPTRQGHFI VMEGAPGARMKPNHNRHVFLFRNHLVTCRPRDSRT | 5940 |
| Qy | 5941 | DTVSVFRNMKLSISIDNDQVGDDEDAFVEMQWREDSVRKYLLOASTAITIKSSWVKEIC | 6000 |
| Db | 5941 | DTVSVFRNMKLSISIDNDQVGDDEDAFVEMQWREDSVRKYLLOASTAITIKSSWVKEIC | 6000 |
| Qy | 6001 | GIQORLALFVWRPDPFEBELADCTAEALGETVKLACRVGTGTPKPIVSWYKDGKAVQVDPFH | 6060 |
| Db | 6001 | GIQORLALFVWRPDPFEBELADCTAEALGETVKLACRVGTGTPKPIVSWYKDGKAVQVDPFH | 6060 |
| Qy | 6061 | ILIEDPDGSCALILDSILTGVDGSGQYMCFAAAGNCSTLKGILVOPFRFVKNKTRASPFF | 6120 |
| Db | 6061 | ILIEDPDGSCALILDSILTGVDGSGQYMCFAAAGNCSTLKGILVOPFRFVKNKTRASPFF | 6120 |

QY 6121 EGEDAQTCTTIEGAPYPOIRYKDGALLTTGNKQTLSEPRSGLLVLVIRAASKEDLGLY 6180
DB 6121 EGEDAQTCTTIEGAPYPOIRYKDGALLTTGNKQTLSEPRSGLLVLVIRAASKEDLGLY 6180
QY 6181 ECELVNLGARSASAEALRIQSPMLQAQOECHREQLVAAVEDTTLERADQEVTSVLKELLG 6240
DB 6181 ECELVNLGARSASAEALRIQSPMLQAQOECHREQLVAAVEDTTLERADQEVTSVLKELLG 6240
QY 6241 PKAPGSPGDTLGTGCPGPRGAPALQETGSPQPVVTGTSAPAVPRPVQPLLHEGPEQBE 6300
DB 6241 PKAPGSPGDTLGTGCPGPRGAPALQETGSPQPVVTGTSAPAVPRPVQPLLHEGPEQBE 6300
QY 6301 A1ARAQEWTVPIRMGAAMPAGAGTGELLWDVHSHVRETTORTVYQAIIDHTARPPSMQ 6360
DB 6301 A1ARAQEWTVPIRMGAAMPAGAGTGELLWDVHSHVRETTORTVYQAIIDHTARPPSMQ 6360
QY 6361 VTIEDVQAQTGTAQFAEAIIEGDPQPSVTWKDQSVQLVDSSTRLSQOQEGTYSVLVSHVA 6420
DB 6361 VTIEDVQAQTGTAQFAEAIIEGDPQPSVTWKDQSVQLVDSSTRLSQOQEGTYSVLVSHVA 6420
QY 6421 SKDAGVYTCIAQNTGGQVLCFAELVLLVGGDNEPDEKOSHRKLLHSFYEKVEE1GRGVFG 6480
DB 6421 SKDAGVYTCIAQNTGGQVLCFAELVLLVGGDNEPDEKOSHRKLLHSFYEKVEE1GRGVFG 6480
QY 6481 FVKRVQHKGNKILCAAKFIPLRSRTAQAYRERDILAAALSHPLVTGLDQFETRKTLILI 6540
DB 6481 FVKRVQHKGNKILCAAKFIPLRSRTAQAYRERDILAAALSHPLVTGLDQFETRKTLILI 6540
QY 6541 LELCSSBELLRLYKGVVTAEVKVIYIQQLVEGLHYLHSHGVUHLDIKPSNIMLWHPAR 6600
DB 6541 LELCSSBELLRLYKGVVTAEVKVIYIQQLVEGLHYLHSHGVUHLDIKPSNIMLWHPAR 6600
QY 6601 EDIKICDPGFAQNTTPAELQSVGSPEFVSPEI1QQNPVSEASDIWANGVVISLSLTC 6660
DB 6601 EDIKICDPGFAQNTTPAELQSVGSPEFVSPEI1QQNPVSEASDIWANGVVISLSLTC 6660
QY 6661 SPFAGESDRATLLNVLEGRVSWSPMAHLSEDAKDFTKATLQAPQAPSAOCLSHPW 6720
DB 6661 SPFAGESDRATLLNVLEGRVSWSPMAHLSEDAKDFTKATLQAPQAPSAOCLSHPW 6720
QY 6721 FLKSNPAEAFINTKQJFLKLLARSWORSLSWYKSLVWES1PELLARGPDSLSGLVAR 6780
DB 6721 FLKSNPAEAFINTKQJFLKLLARSWORSLSWYKSLVWES1PELLARGPDSLSGLVAR 6780
QY 6781 HLCDRTGSS 6840
DB 6781 HLCDRTGSS 6840
QY 6841 RSTAPAPAPGAGPAGPAAQGCVPVRSVIRSLFYHQAGESPENGALAPGSRHRPARRH 6900
DB 6841 RSTAPAPAPGAGPAGPAAQGCVPVRSVIRSLFYHQAGESPENGALAPGSRHRPARRH 6900
QY 6901 LLKGGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSFETALRLPASGTHLAPGH 6960
DB 6901 LLKGGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSFETALRLPASGTHLAPGH 6960
QY 6961 SHSLEHDSFSTRPSSEACGSAQRLPSAPSGAP1RDMGHPPQGSKQLPSTGCGHPTAQPE 7020
DB 6961 SHSLEHDSFSTRPSSEACGSAQRLPSAPSGAP1RDMGHPPQGSKQLPSTGCGHPTAQPE 7020
QY 7021 RPSDPSWGQAPFCHPQGSAPQGCSPHAPVAPCPGSPPGCSKEAPLVPSSPFLGQ 7080
DB 7021 RPSDPSWGQAPFCHPQGSAPQGCSPHAPVAPCPGSPPGCSKEAPLVPSSPFLGQ 7080
QY 7081 PQAPPAPAKAPPLDKNKPGD1SLPGRPKGPCSSPGSASQASSSQVSSLRVSSQVGT 7140
DB 7081 PQAPPAPAKAPPLDKNKPGD1SLPGRPKGPCSSPGSASQASSSQVSSLRVSSQVGT 7140
QY 7141 EPGSLDAEGWTOEAED1SDSTPTLQROEQATWKRS1LGGRGYAGVAGTG1FAFGDA 7200
DB 7141 EPGSLDAEGWTOEAED1SDSTPTLQROEQATWKRS1LGGRGYAGVAGTG1FAFGDA 7200
QY 7201 GGMLQGGPMMARIAWAVSQSEEEQBEARAESQSEEQARAEASPLPQVSARPVEVGRA 7260

DB 7201 GGMLQGGPMMARIAWAVSQSEEEQBEARAESQSEEQARAEASPLPQVSARPVEVGRA 7260
QY 7261 PTRSSPEPTWEDIGQVSLVQIRDLSGDAEADT1SLDISEYDPAYLNLSDLYDKYLPF 7320
DB 7261 PTRSSPEPTWEDIGQVSLVQIRDLSGDAEADT1SLDISEYDPAYLNLSDLYDKYLPF 7320
QY 7321 EFMIFKVPKSAQPPPPSPMAEEELAEPEPTWMPGELGPHAGLEITEESDNDALLAE 7380
DB 7321 EFMIFKVPKSAQPPPPSPMAEEELAEPEPTWMPGELGPHAGLEITEESDNDALLAE 7380
QY 7381 AAVGRKRKWSFSPRSILFHPGSHLPLDEPAELGLERVKASVEHISRL1KGRPEGLKEG 7440
DB 7381 AAVGRKRKWSFSPRSILFHPGSHLPLDEPAELGLERVKASVEHISRL1KGRPEGLKEG 7440
QY 7441 PRKXEGGLASFR1SLGLKSWDRAPTF1RELSDET1V1LQSVTLACQVSAQPAQAATWSKDG 7500
DB 7441 PRKXEGGLASFR1SLGLKSWDRAPTF1RELSDET1V1LQSVTLACQVSAQPAQAATWSKDG 7500
QY 7501 APLESSSRVLI1SATLKNFOLL1L1VVAEDLGVTCVSNALGT1TTTGV1LKAERPSS 7560
DB 7501 APLESSSRVLI1SATLKNFOLL1L1VVAEDLGVTCVSNALGT1TTTGV1LKAERPSS 7560
QY 7561 PCPDIGEVADGVLLVWKPVESYGPVTY1VQCSLEGGSWTTLASDI1FDCCYLT1SKLSRG 7620
DB 7561 PCPDIGEVADGVLLVWKPVESYGPVTY1VQCSLEGGSWTTLASDI1FDCCYLT1SKLSRG 7620
QY 7621 TYTERTACYSKAGMPYSSPSEQVLLGGPSHL1ASEEESQGRSAQPLPSTKTFAFOTQIOR 7680
DB 7621 TYTERTACYSKAGMPYSSPSEQVLLGGPSHL1ASEEESQGRSAQPLPSTKTFAFOTQIOR 7680
QY 7681 GRFVVRQCEWAKSAGSALA1PYHPKDKTAV1REYEALKGLRHPHLAQLHAAYLSPRH 7740
DB 7681 GRFVVRQCEWAKSAGSALA1PYHPKDKTAV1REYEALKGLRHPHLAQLHAAYLSPRH 7740
QY 7741 LV1L1ELCSGP1LLPCLAEASYSSEVKDY1WQVLSATQY1LHNOH1LHLD1RSENMIIT 7800
DB 7741 LV1L1ELCSGP1LLPCLAEASYSSEVKDY1WQVLSATQY1LHNOH1LHLD1RSENMIIT 7800
QY 7801 EYNLLKVVDL1GNAQSLSQKVLPSDKFDY1L1ETMAPELLEGOGAVPQTD1W1GVTAF1M 7860
DB 7801 EYNLLKVVDL1GNAQSLSQKVLPSDKFDY1L1ETMAPELLEGOGAVPQTD1W1GVTAF1M 7860
QY 7861 LSAEYVPSSEGA1DLQ1RGLKGLVRLSR1CYAG1SGGAV1FLST1LCAQ1PWGRPCASSCLQ 7920
DB 7861 LSAEYVPSSEGA1DLQ1RGLKGLVRLSR1CYAG1SGGAV1FLST1LCAQ1PWGRPCASSCLQ 7920
QY 7921 CPWLTEEGPACSRPA1PVTPTAR1RVFVNR1KRALL1YKRN1LAQVR 7968
DB 7921 CPWLTEEGPACSRPA1PVTPTAR1RVFVNR1KRALL1YKRN1LAQVR 7968
RESULT 2
ID ABG76186
XX ABG76186 standard; protein; 2630 AA.
AC ABG76186;
XX
DT 09-MAY-2003 (first entry)
XX
DE Human serine/threonine or protein kinase 59079.
XX
KW Human; enzyme; serine/threonine kinase; protein kinase; 59079;
KW cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma; immunogen;
KW blood platelet disorder; thrombocytopenia; leukaemia; Hodgkin's disease;
KW haemolytic anaemia; cellular proliferative disorder; cancer;
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.
XX
OS Homo sapiens.
XX

| | | | | | | |
|----------------------------|---|------|------|---|--------------------------------------|------|
| PN | US2002168742-A1. | QY | 5579 | SRPSSSIQESSSESDGDARGEI | FDIYVVTADYPLGABQDAITLREGQYVEVLDAAHPL | 5638 |
| XX | | DB | 241 | SRPSSSIQESSSESDGDARGEI | FDIYVVTADYPLGABQDAITLREGQYVEVLDAAHPL | 300 |
| PD | 14-NOV-2002. | | | | | |
| XX | | | | | | |
| PF | 15-FEB-2002; 2002US-00077130. | QY | 5639 | RWLVRTKPTKSSPSRQGWSPAYLDRRLKLSPEWGAAEAPPEPGEAVSDEYKARLSSVI | 5698 | |
| XX | | DB | 301 | RWLVRTKPTKSSPSRQGWSPAYLDRRLKLSPEWGAAEAPPEPGEAVSDEYKARLSSVI | 360 | |
| XX | | | | | | |
| PR | 15-FEB-2001; 2001US-0269201P. | QY | 5699 | QELLSSEQAFFVEELQFLQSHHLQHLERCPHPVPIAVAGQKAVIFRNVRDIGRHSFSLQEL | 5758 | |
| XX | (MILL-) MILLENNIUM PHARM INC. | DB | 361 | QELLSSEQAFFVEELQFLQSHHLQHLERCPHPVPIAVAGQKAVIFRNVRDIGRHSFSLQEL | 420 | |
| PA | Kapeller-Libermann R, Acton SL; | | | | | |
| PI | | | | | | |
| XX | | QY | 5759 | QOCDDDDVAMCFIKNOAAFEQYLEFLVGEVQAESVVVSTAIOEFYKKYAEALLAGDPS | 5818 | |
| DR | WPI; 2003-298729/29. | DB | 421 | QOCDDDDVAMCFIKNOAAFEQYLEFLVGEVQAESVVVSTAIOEFYKKYAEALLAGDPS | 480 | |
| XX | N-PSDB; ABX11641. | | | | | |
| XX | | QY | 5819 | QPPPPLOHYLEQPVVERVORVQALLKELIRKARNRQNCALLEQAYAVVSALPORAENKL | 5878 | |
| PT | Novel isolated human protein kinase, designated 59079 or 12599 | DB | 481 | QPPPPLOHYLEQPVVERVORVQALLKELIRKARNRQNCALLEQAYAVVSALPORAENKL | 540 | |
| PT | polypeptide, useful as diagnostic and therapeutic agents for preventing | | | | | |
| PT | cardiovascular diseases, proliferative disorders, and protein kinase | | | | | |
| PT | disorders. | | | | | |
| PS | Claim 8; Page 48-54; 119pp; English. | QY | 5879 | HVSLMENYPGTLEALGEPHQGHFIVWEGAPGARMKGNHRHVFLEFRNHLVICPDRDS | 5938 | |
| XX | | DB | 541 | HVSLMENYPGTLEALGEPHQGHFIVWEGAPGARMKGNHRHVFLEFRNHLVICPDRDS | 600 | |
| CC | The invention relates to an isolated human serine/threonine or protein | QY | 5939 | RTDTVSVFRNMKLSIDLDNDQVEGDDRAFEVWQEREDSVKYLQARTAIKSSWVKE | 5998 | |
| CC | kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule | DB | 601 | RTDTVSVFRNMKLSIDLDNDQVEGDDRAFEVWQEREDSVKYLQARTAIKSSWVKE | 660 | |
| CC | comprising at least 85% identity to the nucleic acids appearing as | | | | | |
| CC | ABX11641 and ABX11642 or their complement, a naturally occurring variant | QY | 5999 | ICGICQRLALPWRRPDPFEELADCTAELGTGTVKACRVGTGTPKPVISWYKDGKAVQVDP | 6058 | |
| CC | of the kinases or their fragments. Also included are a non-human host | DB | 661 | ICGICQRLALPWRRPDPFEELADCTAELGTGTVKACRVGTGTPKPVISWYKDGKAVQVDP | 720 | |
| CC | cell containing the nucleic acids, an antibody specific for the proteins, | QY | 6059 | HHILIEDPGSCALILDLSLTGVDGQYMCFAASAGNCSTLGKILVQVPPFRVNVKVRASP | 6118 | |
| CC | identifying a compound which binds to the kinase (by contacting the | DB | 721 | HHILIEDPGSCALILDLSLTGVDGQYMCFAASAGNCSTLGKILVQVPPFRVNVKVRASP | 780 | |
| CC | kinase or a cell expressing the kinase with a test compound and | | | | | |
| CC | determining whether the kinase binds to the test compound) and modulating | QY | 6119 | FVEGEDAQFTCTIEGAPYQIRWYKDGALLTGKNTGNKFTLSEPRSGLLVIVIRAAKEDLG | 6178 | |
| CC | the activity of kinase using the identified compound. The kinases and | DB | 781 | FVEGEDAQFTCTIEGAPYQIRWYKDGALLTGKNTGNKFTLSEPRSGLLVIVIRAAKEDLG | 840 | |
| CC | their encoding nucleic acids are useful as diagnostic and therapeutic | QY | 6179 | LYECELNVRLGSRASAEIRIQSPMLQAOEQCHREQLVAAYVEDTTLERADQVTSVLKRL | 6238 | |
| CC | agents for preventing a disease or condition associated with an aberrant | DB | 841 | LYECELNVRLGSRASAEIRIQSPMLQAOEQCHREQLVAAYVEDTTLERADQVTSVLKRL | 900 | |
| CC | or unwanted 59079 or 12599 activity in a subject, including | QY | 6239 | LGPKAPGSTGDLTGPGCPRGAPALQETGSPVTVGTSEAPVPPRVQPLHLHGPPQE | 6298 | |
| CC | cardiovascular diseases such as heart failure, and myocardial infarction; | DB | 901 | LGPKAPGSTGDLTGPGCPRGAPALQETGSPVTVGTSEAPVPPRVQPLHLHGPPQE | 960 | |
| CC | disorders involving blood vessels such as atherosclerosis, and Kaposi's | QY | 6299 | PEAIRAQEQWTVPIRMEGAAMPAGAGTBELLWDVHSHVVRETTQRTYTYTQAIDHTARPPS | 6358 | |
| CC | sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia, | DB | 961 | PEAIRAQEQWTVPIRMEGAAMPAGAGTBELLWDVHSHVVRETTQRTYTYTQAIDHTARPPS | 1020 | |
| CC | Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders | | | | | |
| CC | such as cancer; and protein kinase disorders such as autoimmune | QY | 6359 | MQVTIEDVQAGTGTGAQFEAIEGDPQPSVTWYKDSVOLVDSTRLSQOQEGTTSVLVRH | 6418 | |
| CC | disorders, diabetes mellitus, psoriasis, inflammatory bowel disease, | DB | 1021 | MQVTIEDVQAGTGTGAQFEAIEGDPQPSVTWYKDSVOLVDSTRLSQOQEGTTSVLVRH | 1080 | |
| CC | rheumatoid arthritis, and multiple sclerosis (many examples of diseases | QY | 6419 | VASKDAGVYTCIAQNTGGQVLCBAELLVLGGDNPDSEKQSHRRKLSHFYEVKEIGRGV | 6478 | |
| CC | and disorders are included in the specification). The kinases, their | DB | 1081 | VASKDAGVYTCIAQNTGGQVLCBAELLVLGGDNPDSEKQSHRRKLSHFYEVKEIGRGV | 1140 | |
| CC | encoding nucleic acids and antibodies are useful in screening assays, | QY | 6479 | FGFVKRVQHGKNIKCAAKFIPLSRTRAQAYRERDILAAALSHPLVTGLDQFTRKTLI | 6538 | |
| CC | detection assays (e.g. forensic biology), and predictive medicine (e.g. | DB | 1141 | FGFVKRVQHGKNIKCAAKFIPLSRTRAQAYRERDILAAALSHPLVTGLDQFTRKTLI | 1200 | |
| CC | diagnostic assays, prognostic assays, and monitoring clinical trials and | QY | 6539 | LILELCSSEELLDRLYRGVVTAEVKYIQQLVEGLHYLHSHGVHLHDIKPSNIMVHP | 6598 | |
| CC | pharmacogenomics). The kinases and their encoding nucleic acids are | DB | 1201 | LILELCSSEELLDRLYRGVVTAEVKYIQQLVEGLHYLHSHGVHLHDIKPSNIMVHP | 1260 | |
| CC | useful as query sequences to perform a search against public databases to | QY | 6599 | AREDIKICDFGFAQNTIPAELOFQSGSPFVSPPEIQQNPVSEASDIWAMGVTSYLSLT | 6658 | |
| CC | identify other family members or related sequences. The present sequence | DB | 1261 | AREDIKICDFGFAQNTIPAELOFQSGSPFVSPPEIQQNPVSEASDIWAMGVTSYLSLT | 1320 | |
| CC | represents kinase 59079 | QY | 6659 | CSSPAGESDRAITLLNVLEGRVSNSSPMAHLSEDAKDFIKATLQRAFPARPSAAQCCLSH | 6718 | |
| XX | | | | | | |
| SQ | Sequence 2630 AA; | | | | | |
| Query Match | 33.2%; Score 13710; DB 6; Length 2630; | | | | | |
| Best Local Similarity | 100.0%; Pred. No. 0; | | | | | |
| Matches 2630; Conservative | 0; Mismatches 0; Indels 0; Gaps 0; | | | | | |
| QY | 5339 MGAAYSSARLLVRGDEPEEPASDVHQLVPPMLERFTPKVKKGSSITFSVKVEGRP | 5398 | | | | |
| DB | 1 MGAAYSSARLLVRGDEPEEPASDVHQLVPPMLERFTPKVKKGSSITFSVKVEGRP | 60 | | | | |
| QY | 5399 VPTVHVLREEARGVLTGPDTPGYTVASSAQHSLVLLDVGRHQHGYTTCIASNAAGQA | 5458 | | | | |
| DB | 61 VPTVHVLREEARGVLTGPDTPGYTVASSAQHSLVLLDVGRHQHGYTTCIASNAAGQA | 120 | | | | |
| QY | 5459 LCSASLHVSGLPKVQEOKVKEALISITFLOQTQAIQAQGLTASPADLGGQRKEPPLAA | 5518 | | | | |
| DB | 121 LCSASLHVSGLPKVQEOKVKEALISITFLOQTQAIQAQGLTASPADLGGQRKEPPLAA | 180 | | | | |
| QY | 5519 KEALGHLSLAEVGTBEFLQKTSQITEMVSAKICTAKLQVPFGDSDEDSKTPSASPRGR | 5578 | | | | |
| DB | 181 KEALGHLSLAEVGTBEFLQKTSQITEMVSAKICTAKLQVPFGDSDEDSKTPSASPRGR | 240 | | | | |

| | | | |
|----|------|--|------|
| QY | 5339 | MGAAYSSARLLVRGPDPEEKPSADVHEQLVPPRMLERFTPKVKKGSSITFSVKVEGRP | 5398 |
| DB | 1 | MGAAYSSARLLVRGPDPEEKPSADVHEQLVPPRMLERFTPKVKKGSSITFSVKVEGRP | 60 |
| QY | 5399 | VPTVHVLREEAERGVLWIGDTPGYTVASSAQHSLVLLDVGRHQHGYTTCIASNAAGQA | 5458 |
| DB | 61 | VPTVHVLREEAERGVLWIGDTPGYTVASSAQHSLVLLDVGRHQHGYTTCIASNAAGQA | 120 |
| QY | 5459 | LCSASLVSGLPKVEEQEKYKEALISTFLOTTQAIQAQGLTASPADLGGQKEPLAA | 5518 |
| DB | 121 | LCSASLVSGLPKVEEQEKYKEALISTFLOTTQAIQAQGLTASPADLGGQKEPLAA | 180 |
| QY | 5519 | KEALGHLSLAERGTEEFQKLQTSITMWSAKITQAKLVQPGGDSDBDSTKTPSAPRHGR | 5578 |
| DB | 181 | KEALGHLSLAERGTEEFQKLQTSITMWSAKITQAKLVQPGGDSDBDSTKTPSAPRHGR | 240 |

Query Match 33.2%; Score 13710; DB 6; Length 2630;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2630; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sequence 2630 AA;

1321 CSSPAGESDRATLNTLEGRVSSPMAAHLSDAKDFIKATLQAPQAPRAQAQCLSH 1380
6719 PWF1KSPAEBAHINTKOLKFLARSRWORSLSYKSIILVMSRIPELLRGPDPSPSLGV 6778
1381 PWF1KSPAEBAHINTKOLKFLARSRWORSLSYKSIILVMSRIPELLRGPDPSPSLGV 1440
6779 ARHLCRDTGSSSSSSSSDNELAPFAKASLPSPSVTHSPLHPRGFLRPSASLPEAEAE 6938
1441 ARHLCRDTGSSSSSSSSDNELAPFAKASLPSPSVTHSPLHPRGFLRPSASLPEAEAE 1500
6939 SERSTEAPAPASPEGACPPAOCVPRHSVIRSLFYHQGESPEHGALAPGRRHPARR 6998
1501 SERSTEAPAPASPEGACPPAOCVPRHSVIRSLFYHQGESPEHGALAPGRRHPARR 1560
6899 RHLLKGGYIAGALPGLREPLMEHVRVLEBEAREEQATLLAKAPSFETALRLPASGTHLAP 6958
1561 RHLLKGGYIAGALPGLREPLMEHVRVLEBEAREEQATLLAKAPSFETALRLPASGTHLAP 1620
6959 GHSHSLHDSSTPRPSEACGEAQRULPSAPSGGAPIRDMGHPOGSKQLPSTGCHPGTAQ 7018
1621 GHSHSLHDSSTPRPSEACGEAQRULPSAPSGGAPIRDMGHPOGSKQLPSTGCHPGTAQ 1680
7019 PERSPSPWQOPAFCHPKQGSAPQSGSPHPAVAPCPGSPGPPGCKEAPLVPSSPFL 7078
1681 PERSPSPWQOPAFCHPKQGSAPQSGSPHPAVAPCPGSPGPPGCKEAPLVPSSPFL 1740
7079 GQOAPAPAKASPLDCKMGPDISLPGPKPGCSPPSGASQASSQVSSLRVGSQV 7138
1741 GQOAPAPAKASPLDCKMGPDISLPGPKPGCSPPSGASQASSQVSSLRVGSQV 1800
7139 GTEPGPSLDAGWTQAEADLSDSTPTLQRPQEQATMRKFSLGGGGYAGVAGYGTAFGG 7198
1801 GTEPGPSLDAGWTQAEADLSDSTPTLQRPQEQATMRKFSLGGGGYAGVAGYGTAFGG 1860
7199 DAGMLGQGPWARIAMAVSQSEEEBEAREAEQSEEEQOAPAESELPQVSRAPVPEVG 7258
1861 DAGMLGQGPWARIAMAVSQSEEEBEAREAEQSEEEQOAPAESELPQVSRAPVPEVG 1920
7259 RAPTRSSPEPTWEDIGQVSLVQIRLSDGDAEADTTSLDISEVDPRAYLNSLDYDKYL 7318
1921 RAPTRSSPEPTWEDIGQVSLVQIRLSDGDAEADTTSLDISEVDPRAYLNSLDYDKYL 1980
7319 PFEMIFRKPQKQPEPPSPMAEELAEFPETWMPGELPHAGLEITEESDDVDALL 7378
1981 PFEMIFRKPQKQPEPPSPMAEELAEFPETWMPGELPHAGLEITEESDDVDALL 2040
7379 ABAVGRKRKXWSSPSRSLFHPGRHLPLDEPAELGLRERVKASVEHISRLKGRPEGLEK 7438
2041 ABAVGRKRKXWSSPSRSLFHPGRHLPLDEPAELGLRERVKASVEHISRLKGRPEGLEK 2100
7439 EGPPRKPGLASFRSLGKSDRAPTFRLRLSDETVLQSVTLACQVSAQPAQAATWSK 7498
2101 EGPPRKPGLASFRSLGKSDRAPTFRLRLSDETVLQSVTLACQVSAQPAQAATWSK 2160
7499 DGAPLESSRVLISATLKNFOLLITLVVAEDLGVYTCVSNALGTVTITGVLEKABRPS 7558
2161 DGAPLESSRVLISATLKNFOLLITLVVAEDLGVYTCVSNALGTVTITGVLEKABRPS 2220
7559 SSPCPDICEVADGVLLVWKVSVESGPTVYVQCSLEGSSWTTLASDIFDCCYLTSLSR 7618
2221 SSPCPDICEVADGVLLVWKVSVESGPTVYVQCSLEGSSWTTLASDIFDCCYLTSLSR 2280
7619 GGTYTFRTACVSKAGMGVSPSEQVLLGGPSHLASEESQGRSAQPLPSTKTFAFQTQI 7678
2281 GGTYTFRTACVSKAGMGVSPSEQVLLGGPSHLASEESQGRSAQPLPSTKTFAFQTQI 2340
7679 QGRFSVVRQCKEASGALAAKIIPYHPKDKTAVLREYKALKGRHPLHQAALHAALYSP 7738
2341 QGRFSVVRQCKEASGALAAKIIPYHPKDKTAVLREYKALKGRHPLHQAALHAALYSP 2400
7739 RHLVLIELCSPGPELLCLAEASVSESEVKDYLWQMLSATQYLNHQLHLDLRSENNI 7798

2401 RHLVLIELCSPGPELLCLAEASVSESEVKDYLWQMLSATQYLNHQLHLDLRSENNI 2460
7799 ITEYNLLKVVLDLGNASQSLSQEKVLPDSKDFKDYLETMAPELLEGQGVAPQOTDIWALGVTA 7858
2461 ITEYNLLKVVLDLGNASQSLSQEKVLPDSKDFKDYLETMAPELLEGQGVAPQOTDIWALGVTA 2520
7859 IMLSAEYFVSSSGARDLQRLKGLVRLSRVRCVAGLSGGAVAFRLSTICAOPHGRPCASSC 7918
2521 IMLSAEYFVSSSGARDLQRLKGLVRLSRVRCVAGLSGGAVAFRLSTICAOPHGRPCASSC 2580
7919 LQCPMLTEGPACSRPAPVTFPTALRLVFRNRREKRALLYKRNHLAQRV 7968
2581 LQCPMLTEGPACSRPAPVTFPTALRLVFRNRREKRALLYKRNHLAQRV 2630

RESULT 3
AAB30569
ID AAB30569 standard; protein; 2596 AA.
XX
AC AAB30569;
XX
DT 19-MAR-2001 (first entry)
XX
DE A splice variant of a signal transduction polypeptide.
XX
KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
congestive heart failure; dilated congestive cardiomyopathy;
hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
mitral valve disease; aortic valve disease; tricuspid valve disease;
myocardial infarction; cardiac arrhythmia; arteriosclerosis;
atherosclerosis; cardiac tumour; microbial infection; splice variant.
XX
OS Homo sapiens.
XX
FH Key
FT Domain 325..504 Location/Qualifiers
FT Domain /note= "guanine nucleotide exchange factor domain"
FT Domain 1094..1351
FT Domain /note= "kinase domain"
FT Domain 2301..2553
FT Domain /note= "kinase domain"
XX
FN WO200063381-A1.
XX
XX 26-OCT-2000.
XX
XX 11-APR-2000; 2000WO-US009488.
XX
XX 16-APR-1999; 99US-0129553P.
XX
XX (SCIO-) SCIOS INC.
XX
XX Zeng W, Stanton L, Kong H;
XX
XX WPI; 2001-007013/01.
XX
XX N-PSDB; AAC62287.
XX
XX Novel h19G5 polypeptides capable of regulating signal transduction and
exhibiting kinase activity useful for identifying antibodies to treat
cardiac diseases, and additional mediators of signal transduction.
XX
XX Claim 1; Page 68-74; 81pp; English.
XX
XX The present sequence represents a splice variant of human in signal
transduction polypeptide. The polypeptide is designated H19G5. The
protein is capable of regulating signal transduction and exhibits kinase
activity. The H19G5 transcript is expressed in the heart. H19G5
polypeptides and polynucleotides are useful for preventing or treating a
cardiac disease, such as congestive heart failure, dilated congestive
cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
mitral valve disease, aortic valve disease or tricuspid valve disease,
angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
arterial or renovascular hypertension, arteriosclerosis, atherosclerosis

| | |
|----|--|
| CC | and cardiac tumours in humans. The polypeptide is also useful for |
| CC | detecting the expression of a protein capable of regulating signal |
| CC | transduction or the expression of a protein capable of acting as a donor |
| CC | or acceptor molecule of a phosphate group. The monoclonal antibodies can |
| CC | be used as probes for detecting discrete antigens expressed by tissue or |
| CC | cell samples, and therefore used in humans for localization and |
| CC | monitoring of microbial infection |
| xx | |
| SQ | Sequence 2596 AA; |
| | Query Match 32.8%; Score 13528; DB 4; Length 2596; |
| | Best Local Similarity 100.0%; Pred. No. 0; |
| | Matches 2595; Conservative 0; Mismatches 1; Indels 0; Gaps 0; |
| QY | 5373 MLERFTPKKVKKSSITFSVKEGRPVPTVHVLREEAERGVLWIGDTPGYTVASSAQQH 5432 |
| DB | 1 MLERFTPKKVKKSSITFSVKEGRPVPTVHVLREEAERGVLWIGDTPGYTVASSAQQH 60 |
| QY | 5433 SLVLLDVGRHQGTCTIASNAAGALCSALHVSGLPKVEEQBKVKREALISTFLOQTQ 5492 |
| DB | 61 SLVLLDVGRHQGTCTIASNAAGALCSALHVSGLPKVEEQBKVKREALISTFLOQTQ 120 |
| QY | 5493 AISAQGLTASFADLGGORKEEPLAAKEALGHLSLAEVGTETFLQKLTSTITWVSAKIT 5552 |
| DB | 121 AISAQGLTASFADLGGORKEEPLAAKEALGHLSLAEVGTETFLQKLTSTITWVSAKIT 180 |
| QY | 5553 QAKLVQPGDSEDESKTPSASPRHGRSRPSSIOESSSESEDGDARGEIIDIYVVTADYL 5612 |
| DB | 181 QAKLVQPGDSEDESKTPSASPRHGRSRPSSIOESSSESEDGDARGEIIDIYVVTADYL 240 |
| QY | 5613 PLGAEQDAITUREGOYVEVLDAAPLAWLVRTKTSKSPRGWVSPAYLDRLKLSPEW 5672 |
| DB | 241 PLGAEQDAITUREGOYVEVLDAAPLAWLVRTKTSKSPRGWVSPAYLDRLKLSPEW 300 |
| QY | 5673 GAAEAPFPFGGAVSEDEYKARLSSVIOELLSSEQAFVEELQFLQSHLQHLERCPHVPIA 5732 |
| DB | 301 GAAEAPFPFGGAVSEDEYKARLSSVIOELLSSEQAFVEELQFLQSHLQHLERCPHVPIA 360 |
| QY | 5733 VAGKAVIFRNVRDIGHFSSFLQELQCDTDDVAMCFIKNOAAFEQYLEFVLGVQAE 5792 |
| DB | 361 VAGKAVIFRNVRDIGHFSSFLQELQCDTDDVAMCFIKNOAAFEQYLEFVLGVQAE 420 |
| QY | 5793 SVVYSTAIQEPYKKAEEALLAGDPSPPPPLQCHYLEQVVERVQRYQALLKELIRNKAR 5852 |
| DB | 421 SVVYSTAIQEPYKKAEEALLAGDPSPPPPLQCHYLEQVVERVQRYQALLKELIRNKAR 480 |
| QY | 5853 NRQNCALLEQAYAVVSALPQAEKHLVSLMENYPTGLEALGFPIROGHFIVNEGAPGAR 5912 |
| DB | 481 NRQNCALLEQAYAVVSALPQAEKHLVSLMENYPTGLEALGFPIROGHFIVNEGAPGAR 540 |
| QY | 5913 MPWKGNHRHVFLEFNHLVICKPRDSDRTDTSVYVFRNMKLSIDLDQVGGDDRAFEVW 5972 |
| DB | 541 MPWKGNHRHVFLEFNHLVICKPRDSDRTDTSVYVFRNMKLSIDLDQVGGDDRAFEVW 600 |
| QY | 5973 QERDSVRKYLQARTAIKSSWVKEICGIQORLALPVRPDPFEBELADCTAELGETVK 6032 |
| DB | 601 QERDSVRKYLQARTAIKSSWVKEICGIQORLALPVRPDPFEBELADCTAELGETVK 660 |
| QY | 6033 LACRVGTGPKPVIISWYKDGKAVQVDPHHILIEDPDGSCALLDLSLTCVDSGQWCFPAASA 6092 |
| DB | 661 LACRVGTGPKPVIISWYKDGKAVQVDPHHILIEDPDGSCALLDLSLTCVDSGQWCFPAASA 720 |
| QY | 6093 AGNCSTLGKILVQVPPFRVKNVRASPEVEGEDAQFTCTIECAPYQPIRWYKDGALLTTGN 6152 |
| DB | 721 AGNCSTLGKILVQVPPFRVKNVRASPEVEGEDAQFTCTIECAPYQPIRWYKDGALLTTGN 780 |
| QY | 6153 KFOQLTSRPRGLVILVIRAAKEDLGLYECELVNRIGLSARASAEIRIQSPMLQAOQCHHR 6212 |
| DB | 781 KFOQLTSRPRGLVILVIRAAKEDLGLYECELVNRIGLSARASAEIRIQSPMLQAOQCHHR 840 |
| QY | 6213 EQLVAAVEDTTLERADQEVTSVLKRLIGPKAPGFSTGDLTGPGFCPRGAPALQETGQPP 6272 |
| DB | 841 EQLVAAVEDTTLERADQEVTSVLKRLIGPKAPGFSTGDLTGPGFCPRGAPALQETGQPP 900 |

| | |
|----|---|
| QY | 6273 VTGTSEAPVPRVPOPLLHEGPEQEPALARAQEWTPVIRMGAAMPAGTGLLWDVH 6332 |
| DB | 901 VTGTSEAPVPRVPOPLLHEGPEQEPALARAQEWTPVIRMGAAMPAGTGLLWDVH 960 |
| QY | 6333 SHVRETTQRTVYQAIIDHTHTARPPSMQVTIEDVQAQTGGTAQFBAIIEGDPQPSVTWYK 6392 |
| DB | 961 SHVRETTQRTVYQAIIDHTHTARPPSMQVTIEDVQAQTGGTAQFBAIIEGDPQPSVTWYK 1020 |
| QY | 6393 DSVQLVDSFRLSQOQEGTYSILVRHVASKDAGVYVTCIAQNTGGVLCCKALLVLGGDNE 6452 |
| DB | 1021 DSVQLVDSFRLSQOQEGTYSILVRHVASKDAGVYVTCIAQNTGGVLCCKALLVLGGDNE 1080 |
| QY | 6453 PDSEKQSHRRKHLHSFYEVKEIGRGVFGFVKRVQHGKNKILCAAKFIPLRSRTRAQAYRE 6512 |
| DB | 1081 PDSEKQSHRRKHLHSFYEVKEIGRGVFGFVKRVQHGKNKILCAAKFIPLRSRTRAQAYRE 1140 |
| QY | 6513 RDIILAAISHPLVTGLLDQFETRKTLLILELCSSEELLDRLYRKGVVTEAEVKYIQQLY 6572 |
| DB | 1141 RDIILAAISHPLVTGLLDQFETRKTLLILELCSSEELLDRLYRKGVVTEAEVKYIQQLY 1200 |
| QY | 6573 EGLHYLHSHGVHLHLDIKPSNLMVHPAREDIKICDFGPAQNITPAELQFSQYGSPEFVSP 6632 |
| DB | 1201 EGLHYLHSHGVHLHLDIKPSNLMVHPAREDIKICDFGPAQNITPAELQFSQYGSPEFVSP 1260 |
| QY | 6633 EIIQONPVSEASDIWAMGVI SYLSITCSSPPAGESDRATLLNVLEGRVSWSSPVAHLSE 6692 |
| DB | 1261 EIIQONPVSEASDIWAMGVI SYLSITCSSPPAGESDRATLLNVLEGRVSWSSPVAHLSE 1320 |
| QY | 6693 DAKDIFKATLQAPQAPRASAQCLSHPMFLKSMAPAEAAHFINTKQLFLLARSQWQSLM 6752 |
| DB | 1321 DAKDIFKATLQAPQAPRASAQCLSHPMFLKSMAPAEAAHFINTKQLFLLARSQWQSLM 1380 |
| QY | 6753 SYKSLVNRSIPELLRGPPDSPSLGVARHLCRDTGGSSSSSSSDNELAPPARAKSLPPS 6812 |
| DB | 1381 SYKSLVNRSIPELLRGPPDSPSLGVARHLCRDTGGSSSSSSSDNELAPPARAKSLPPS 1440 |
| QY | 6813 PVTHSPLHLHPGFLRPSASLPPEAEASERSTEAPASPASPEGAPPAAGCVCPHSHVRS 6872 |
| DB | 1441 PVTHSPLHLHPGFLRPSASLPPEAEASERSTEAPASPASPEGAPPAAGCVCPHSHVRS 1500 |
| QY | 6873 LFYHQAGSPHGHALAPGSRHRPARRRHLKGGYIAGALPGLREPLMEHRVLESEAAARE 6932 |
| DB | 1501 LFYHQAGSPHGHALAPGSRHRPARRRHLKGGYIAGALPGLREPLMEHRVLESEAAARE 1560 |
| QY | 6933 QATLLAKAPSPETALRLPASGTHLAPGSHSHLSHDSSTPRPSSEACGEAQLRPSAPSGG 6992 |
| DB | 1561 QATLLAKAPSPETALRLPASGTHLAPGSHSHLSHDSSTPRPSSEACGEAQLRPSAPSGG 1620 |
| QY | 6993 APIRDMGHPQSKQLPSTGGHPGTAQPERPSPDPSPWGPAPFCHPKGSGAPQEGCSHPHA 7052 |
| DB | 1621 APIRDMGHPQSKQLPSTGGHPGTAQPERPSPDPSPWGPAPFCHPKGSGAPQEGCSHPHA 1680 |
| QY | 7053 VAPCPGSGFPFGSCKEAPLVPSSPFLQOPAPPAKAPAKASPLDLSKMGPGDISLPRKPKG 7112 |
| DB | 1681 VAPCPGSGFPFGSCKEAPLVPSSPFLQOPAPPAKAPAKASPLDLSKMGPGDISLPRKPKG 1740 |
| QY | 7113 PCSSPGSGASQASSQVSSLRVGSSQVGTPEGPPSLDAEGWTQAEADLSDSTPTLQRPQEA 7172 |
| DB | 1741 PCSSPGSGASQASSQVSSLRVGSSQVGTPEGPPSLDAEGWTQAEADLSDSTPTLQRPQEA 1800 |
| QY | 7173 TMRKPSLGGRGYAGVAGYGTFAFGGAGMGLGGPMWARIWAVSSEEEEBQSEARAS 7232 |
| DB | 1801 TMRKPSLGGRGYAGVAGYGTFAFGGAGMGLGGPMWARIWAVSSEEEEBQSEARAS 1860 |
| QY | 7233 QSEEQEAEARASPLPQVSARVPVPGRAPTHSSPEPTWEDIGOVSLVQIRDLSGDAEA 7292 |
| DB | 1861 QSEEQEAEARASPLPQVSARVPVPGRAPTHSSPEPTWEDIGOVSLVQIRDLSGDAEA 1920 |
| QY | 7293 DTISLDISEVDPAYLNLSDLYDIKYLPEFEMI FRKVPKSAQPEPPSPMAEEELAEFPEPT 7352 |
| DB | 1921 DTISLDISEVDPAYLNLSDLYDIKYLPEFEMI FRKVPKSAQPEPPSPMAEEELAEFPEPT 1980 |

| | | | | |
|----------|---|-----------------------------|---|------|
| Qy | 7353 | WPWFGELGPHAGLEITBES | EDVALLAEAAVGRKRWSSPSRSLFHPGRHLPDLDEPAEL | 7412 |
| Db | 1981 | WPWFGELGPHAGLEITBES | EDVALLAEAAVGRKRWSSPSRSLFHPGRHLPDLDEPAEL | 2040 |
| Qy | 7413 | GLRERVKASVEHLSRIILKGR | PEGLEKGGPRKPGLASFELSGIKSWDRAPTFLRELSDE | 7472 |
| Db | 2041 | GLRERVKASVEHLSRIILKGR | PEGLEKGGPRKPGLASFELSGIKSWDRAPTFLRELSDE | 2100 |
| Qy | 7473 | TVVLGQSVTLACQVSAQPA | QAQATWSKDGAFLESSRVLISATLKNFOLLTILVVVAEDLG | 7532 |
| Db | 2101 | TVVLGQSVTLACQVSAQPA | QAQATWSKDGAFLESSRVLISATLKNFOLLTILVVVAEDLG | 2160 |
| Qy | 7533 | VYTCSVSNALGTVTTTGV | LRAERPPSSPCPDGEVYADGVLLVWKPVESYGPVTYIVOC | 7592 |
| Db | 2161 | VYTCSVSNALGTVTTTGV | LRAERPPSSPCPDGEVYADGVLLVWKPVESYGPVTYIVOC | 2220 |
| Qy | 7593 | SLEGGSWTTLASDIFDC | CYLTSKLSRGTYTFTACVSKAGMGPYSSPQVLLGPFSLH | 7652 |
| Db | 2221 | SLEGGSWTTLASDIFDC | CYLTSKLSRGTYTFTACVSKAGMGPYSSPQVLLGPFSLH | 2280 |
| Qy | 7653 | ASESESQGRSAQPLPSKT | FAFQIOGRFSPVVRQCWEKASGRALAAKIIPYHPKDKTA | 7712 |
| Db | 2281 | ASESESQGRSAQPLPSKT | FAFQIOGRFSPVVRQCWEKASGRALAAKIIPYHPKDKTA | 2340 |
| Qy | 7713 | VLREYEAALKGLRHPLA | QLAHAAVLSPRHLVLIELCSGPPELLCLAERASYSESEVKDYL | 7772 |
| Db | 2341 | VLREYEAALKGLRHPLA | QLAHAAVLSPRHLVLIELCSGPPELLCLAERASYSESEVKDYL | 2400 |
| Qy | 7773 | WQMLSATQYLNQHIHL | DLRSENMIITEYNLLKVVDLGNASQSKVLPSPDKFKDYLE | 7832 |
| Db | 2401 | WQMLSATQYLNQHIHL | DLRSENMIITEYNLLKVVDLGNASQSKVLPSPDKFKDYLE | 2460 |
| Qy | 7833 | TMAPELLEGQAVPQTDI | WAFVTFIMLSAEYVPSSEGGARDLQGRKGLVRLSRCYAG | 7892 |
| Db | 2461 | TMAPELLEGQAVPQTDI | WAFVTFIMLSAEYVPSSEGGARDLQGRKGLVRLSRCYAG | 2520 |
| Qy | 7893 | LSGGAVAFRLSTLCAQ | PWGRFCASSCLQCPWLTEGPACSRPAPVTFTPTARLAVFVNRE | 7952 |
| Db | 2521 | LSGGAVAFRLSTLCAQ | PWGRFCASSCLQCPWLTEGPACSRPAPVTFTPTARLAVFVNRE | 2580 |
| Qy | 7953 | KRRALLYKRHNLAQVR | 7968 | |
| Db | 2591 | KRRALLYKRHNLAQVR | 2596 | |
| RESULT 4 | | | | |
| ID | ABP70084 | standard; protein; 4691 AA. | | |
| AC | ABP70084; | | | |
| XX | XX | | | |
| DT | 27-JAN-2003 | (first entry) | | |
| XX | XX | | | |
| DE | Human NOV13a. | | | |
| XX | XX | | | |
| KW | Human; anti-HIV; cytostatic; antidiabetic; antiasthmatic; cachexia; AIDS; | | | |
| KW | antiinflammatory; cardiant; haemostatic; neuroprotective; anorectic; | | | |
| KW | nootropic; immunosuppressive; osteopathic; antiparkinsonian; cancer; | | | |
| KW | antinfertility; cerebroprotective; gene therapy; NOVX; NOV; fertility; | | | |
| KW | metabolic disorder; diabetes; obesity; infectious disease; anorexia; | | | |
| KW | neurodegenerative diseases; Alzheimer's disease; Parkinson's disease; | | | |
| KW | immune disorder; haematopoietic disorder; cardiovascular disorder; | | | |
| KW | bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis; | | | |
| KW | metabolic syndrome X; wasting disorder; cell differentiation; | | | |
| KW | cell proliferation; haematopoiesis; wound healing; angiogenesis. | | | |
| XX | XX | | | |
| OS | Homo sapiens. | | | |
| XX | XX | | | |
| PN | WO200272771-A2. | | | |
| XX | XX | | | |
| PD | 19-SEP-2002. | | | |
| XX | XX | | | |
| PF | 08-MAR-2002; 2002WO-US007288. | | | |

| | | |
|----|-------------------------------|--|
| XX | 08-MAR-2001; 2001US-0274101P. | |
| PR | 08-MAR-2001; 2001US-0274194P. | |
| PR | 08-MAR-2001; 2001US-0274281P. | |
| PR | 08-MAR-2001; 2001US-0274322P. | |
| PR | 09-MAR-2001; 2001US-0274849P. | |
| PR | 12-MAR-2001; 2001US-0275235P. | |
| PR | 13-MAR-2001; 2001US-0275578P. | |
| PR | 13-MAR-2001; 2001US-0275579P. | |
| PR | 13-MAR-2001; 2001US-0275601P. | |
| PR | 14-MAR-2001; 2001US-0276000P. | |
| PR | 16-MAR-2001; 2001US-0276776P. | |
| PR | 19-MAR-2001; 2001US-0276994P. | |
| PR | 20-MAR-2001; 2001US-0277239P. | |
| PR | 20-MAR-2001; 2001US-0277321P. | |
| PR | 20-MAR-2001; 2001US-0277327P. | |
| PR | 20-MAR-2001; 2001US-0277338P. | |
| PR | 21-MAR-2001; 2001US-0277791P. | |
| PR | 22-MAR-2001; 2001US-0277833P. | |
| PR | 23-MAR-2001; 2001US-0278152P. | |
| PR | 26-MAR-2001; 2001US-0278894P. | |
| PR | 27-MAR-2001; 2001US-0278999P. | |
| PR | 27-MAR-2001; 2001US-0279036P. | |
| PR | 28-MAR-2001; 2001US-0279344P. | |
| PR | 30-MAR-2001; 2001US-0279995P. | |
| PR | 30-MAR-2001; 2001US-0280233P. | |
| PR | 02-APR-2001; 2001US-0280802P. | |
| PR | 02-APR-2001; 2001US-0280822P. | |
| PR | 02-APR-2001; 2001US-0280900P. | |
| PR | 04-APR-2001; 2001US-0281194P. | |
| PR | 13-APR-2001; 2001US-0283675P. | |
| PR | 30-APR-2001; 2001US-0287424P. | |
| PR | 02-MAY-2001; 2001US-0288066P. | |
| PR | 03-MAY-2001; 2001US-0288342P. | |
| PR | 03-MAY-2001; 2001US-0288528P. | |
| PR | 15-MAY-2001; 2001US-0291190P. | |
| PR | 16-MAY-2001; 2001US-0291099P. | |
| PR | 16-MAY-2001; 2001US-0291240P. | |
| PR | 31-MAY-2001; 2001US-0294485P. | |
| PR | 31-MAY-2001; 2001US-0294889P. | |
| PR | 31-MAY-2001; 2001US-0294899P. | |
| PR | 18-JUN-2001; 2001US-0299027P. | |
| PR | 19-JUN-2001; 2001US-0299303P. | |
| PR | 19-JUN-2001; 2001US-0299310P. | |
| PR | 10-JUL-2001; 2001US-0304354P. | |
| PR | 31-JUL-2001; 2001US-0309198P. | |
| PR | 16-AUG-2001; 2001US-0312903P. | |
| PR | 10-SEP-2001; 2001US-0318462P. | |
| PR | 12-SEP-2001; 2001US-0318770P. | |
| PR | 27-SEP-2001; 2001US-0325430P. | |
| PR | 27-SEP-2001; 2001US-0325681P. | |
| PR | 18-OCT-2001; 2001US-0330380P. | |
| PR | 31-OCT-2001; 2001US-0335301P. | |
| PR | 14-NOV-2001; 2001US-0332172P. | |
| PR | 14-NOV-2001; 2001US-0332271P. | |
| PR | 14-NOV-2001; 2001US-0332272P. | |
| PR | 14-NOV-2001; 2001US-0333184P. | |
| PR | 21-NOV-2001; 2001US-0333272P. | |
| PR | 03-DEC-2001; 2001US-0332094P. | |
| PR | 03-DEC-2001; 2001US-0337426P. | |
| PR | 04-DEC-2001; 2001US-0338092P. | |
| PR | 04-DEC-2001; 2001US-0337185P. | |
| PR | 03-JAN-2002; 2002US-0345705P. | |
| PR | 08-MAR-2002; 2002US-00093463. | |

(CURA-) CURAGEN CORP.

Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;
PI Boldog FI, Li L, Zerhusen BD, Tchernev VT, Gangoli EA, Vernet CAM;
PI Pena CEA, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;
PI Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;
PI Taupier RJ, Padigaru M, Shenoy SG, Kekuda R, Gusev VY, Pochart PF;
PI Zhong M;

XX WPI; 2002-732824/79.
 DR N-PSDB; ABV99362.
 XX
 PT New NOVX polypeptides and polynucleotides, useful for preventing,
 PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,
 PT Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic
 PT disorders, and asthma.
 PS
 PS Claim 1; Page 137-138; 619pp; English.
 XX
 CC The present invention relates to new isolated proteins (NOVX) and their
 CC coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is
 CC any number from 1 to 48. The NOVX proteins and coding sequences are
 CC useful in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, preferably a NOVX-associated disorder.
 CC The NOVX coding sequences and proteins are useful for treating, diabetes,
 CC preventing or diagnosing diseases such as metabolic disorders, diabetes,
 CC obesity, infectious diseases, anorexia, cancer-associated cachexia,
 CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
 CC disease, immune disorders, hematopoietic disorders, cardiovascular
 CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
 CC disturbances associated with obesity, metabolic syndrome X or wasting
 CC disorders associated with chronic diseases or various cancers. The NOVX
 CC coding sequences and proteins may also be used as targets for the
 CC identification of small molecules that modulate or inhibit e.g.
 CC neurogenesis, cell differentiation, cell proliferation, haematopoiesis,
 CC wound healing and angiogenesis, in gene therapy, in generation of
 CC antibodies that bind immunospecifically to NOVX substances for use in
 CC therapeutic or diagnostic methods
 XX
 SQ Sequence 4691 AA;

Query Match 25.5%; Score 10519.5; DB 5; Length 4691;
 Best Local Similarity 51.1%; Pred. No. 0;
 Matches 2367; Conservative 118; Mismatches 342; Indels 1807; Gaps 66;

3135 LRGSARCOLSHEGHAQLLITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFQALKDL 3194
 541 LRGSARCOLSHEGHAQLLITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFQALKDL 600
 3195 EYLEGGAATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVRLNLRPQDSGRY 3254
 601 EYLEGGAATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVRLNLRPQDSGRY 660
 3255 SCSFGDQTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAAPVWRKGSSETLRD 3314
 661 SCSFGDQTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAAPVWRKGSSETLRD 720
 3315 GDRYCLRDQGMCELRQIRGLAMVDAAEYSCVCGEERTSASLTIRPMAHFTGLRHQESI 3374
 721 GDRYCLRDQGMCELRQIRGLAMVDAAEYSCVCGEERTSASLTIRPMAHFTGLRHQESI 780
 3375 EGATATLRCELSKTAAPVWRKGSSETLRDGRHSLRDQGVAVCELOICGLAVADAGEYSCVC 3434
 781 EGATATLRCELSKTAAPVWRKGSSETLRDGRHSLRDQGVAVCELOICGLAVADAGEYSCVC 840
 3435 GEERTSATLTVKALPAKTEGLRNEEAVEGATAMLCELS KVA PVWRKGPENLRDGDY 3494
 841 GEERTSATLTVKALPAKTEGLRNEEAVEGATAMLCELS KVA PVWRKGPENLRDGDY 900
 3495 ILRQEGTRCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARFIEDVKNQAREGAT 3554
 901 ILRQEGTRCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARFIEDVKNQAREGAT 960
 3555 AVLOCELNSAAPVWRKGSSETLRDGRYSLRQDGTCKELOIRGLAMADTGEYSCVCCQER 3614
 961 AVLOCELNSAAPVWRKGSSETLRDGRYSLRQDGTCKELOIRGLAMADTGEYSCVCCQER 1020
 3615 TSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKVA PVWRKGHETLRDGRHSLRQ 3674
 1021 TSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKVA PVWRKGHETLRDGRHSLRQ 1080
 3675 DGARCELOIRGLAVADAGEYLCMCGKERTSAMLTVRAVPSKFTGLENEEATEGDTATLW 3734
 1081 DGARCELOIRGLAVADAGEYLCMCGKERTSAMLTVRAVPSKFTGLENEEATEGDTATLW 1140
 3735 CELSKAAPVWRKGHETLRDGRHSLRDQGSRCCELOIRGLAVADAGEYSCVCGQERTSAT 3794
 1141 CELSKAAPVWRKGHETLRDGRHSLRDQGSRCCELOIRGLAVADAGEYSCVCGQERTSAT 1200
 3795 LTVRALPARFIEDVKNQAREGATAVLRCELSKTAAPVWRKGSSETLRDGRYSLRQDGTG 3854
 1201 LTVRALPARFIEDVKNQAREGATAVLRCELSKTAAPVWRKGSSETLRDGRYSLRQDGTG 1260
 3855 CELQIHGLSVADTGEYSCVCGQERTSATLTVR ----- 3886
 1261 CELQIHGLSVADTGEYSCVCGQERTSATLTVRALPARFTQDLTKKEASEGATATLQCELS 1320
 3887 ----- 3886
 1321 KVAPVWRKGPETLRDGRYSLKQDGTCELOIRGLHSLVADAGEYSCVCGQERTSATLTVR 1380
 3887 ----- 3886
 1381 ALPARFTEGLRNEEAMSGATATLQCELSKTAAPVWRKGLEALRDGDYKYSLRQDGA VCELO 1440
 3887 ----- 3886
 1441 IHGLAMADNGYSCVCGQERTSATLTVRALPARFIEDVRNKNQKATEGATVTLQCKLRKAAP 1500
 3887 ----- 3886
 1501 VWRKGPENLTKDGRYSLKQDGTSCCELOIRGLVIADAGEYSCICEQERTSATLTVRALPA 1560
 3887 ----- 3886
 1561 RFIEDVRNHEATEGATAVLRCELSKTAAPVWRKGSSETLRDGRYSLRQDGTGRCELQIRGL 1620
 3887 ----- 3886

| | | | | | | | |
|----|------|--|------|----|------|---|------|
| Db | 1621 | AVEDTGEYLCVCGOERTSATLTVRALPARFIDNMNTNOEAREGATATLHCBSKVAVPVMR | 1680 | Db | 2661 | VTPEAGTVSFHLGNHASSAQLTVRAPEVTILEPLQDVQLR | 2701 |
| Qy | 3887 | ----- | 3886 | Qy | 4282 | RWALGGVPIQANEMNDITVEQGTLLHLTLHKVTLEDAGTVSFHVCTCSSEALKVT | 4337 |
| Db | 1681 | KGPETLRDGRHSRLQDGRSCELQIRGLAVVDAGEYSCVCGOERTSATLTVRALPARFIE | 1740 | Db | 2702 | -----GVPIQANEMNDITVEQGTLLHLTLHKVTLEDAGTVSFHVCTCSSEALKVTAVP | 2756 |
| Qy | 3887 | ----- | 3886 | Qy | 4338 | ----- | 4337 |
| Db | 1741 | DVKNOEAREGATAVLOCELSKAAPVWVRKGSSETILRGDRYSLRQDGRCELOIHLGLSVAD | 1800 | Db | 2757 | CLVRGLQNVDFPAGEVATPFCEDGQPSAIAVRDGI FHSIMLSGLGLVADSGTVI FRAGPLV | 2816 |
| Qy | 3887 | ----- | 3886 | Qy | 4338 | ----- | 4337 |
| Db | 1801 | TGEYSCVCGOERTSATLTVRALPARFTQDLTKKEASEGATATLOCELSKVAVPVMKKGPE | 1860 | Db | 2817 | STAKLLIKDPVVVVVSAMODLAVEBGGSALLCOYSRPVQATWKMDEREVHTDGRHVIIE | 2876 |
| Qy | 3887 | ----- | 3886 | Qy | 4338 | ----- | 4337 |
| Db | 1861 | TLRDGRYSLRQDGRCELOIHLGLSVADAGEYSCVCGOERTSATLTVRDCHTLHVMHPY | 1920 | Db | 2877 | QOMNVARLTFRPAIPCDSGIYSCAAAGTRVALLQVQAKNTVVVRGLNENVEALEGGEALPE | 4360 |
| Qy | 3887 | ----- | 3886 | Db | 4361 | COLSQPEVAHAHTWLLDDEPVRTSENAEVVFFENGLRHLHLKXLRPQDSQCRVTFIAGDMV | 2936 |
| Db | 1921 | FQPLKKEBETLIYQIPSPVILFTEGLRNEAMEGATATLOCELSKAAPVWVRKGLE | 1980 | Qy | 2937 | COLSQPEVAHAHTWLLDDEPVRTSENAEVVFFENGLRHLHLKXLRPQDSQCRVTFIAGDMV | 4420 |
| Qy | 3887 | ----- | 3886 | Db | 4421 | TSALFLTVRGDCAVL VQGWLEILEPLKNAAVRAGACAREFTCTLSEAVPVGEASWYINGAA | 2996 |
| Db | 1981 | ALRDGKXYSLRQDGAVCCELQIHLGLAVADNGVYSSIPARFIEDMRNOKATEGATVTLQCKL | 2040 | Qy | 4473 | VQPDSDMTVTADGSHQALLLSAOPHHAGVETFAACRDVAVASARLTVLGLPDPPEDAEVV | 4472 |
| Qy | 3887 | ----- | 3886 | Db | 2997 | TSALFLTVRGDCAVL VQGWLEILEPLKNAAVRAGACAREFTCTLSEAVPVGEASWYINGAA | 3056 |
| Db | 2041 | RKAAPVWVRKGPNTLKDGDRYSLRQDGTSCBLQIRGLVIADAGYSICQOERTSATLTV | 2100 | Qy | 3057 | VQPDSDMTVTADGSHQALLLSAOPHHAGVETFAACRDVAVASARLTVLGLPDPPEDAEVV | 4532 |
| Qy | 3887 | ----- | 3886 | Db | 4533 | ARSSHTVTLSWAAPMSDGGGLCGYRVEVEKEGATQWRLCHELVGPEVCVWGLAPGETY | 3116 |
| Db | 2101 | RALPARFIEDVNRHEATEGATAVLOCELSKAAPVWVRKGSSETLRDGRYSLRQDGRCEL | 2160 | Qy | 3117 | ARSSHTVTLSWAAPMSDGGGLCGYRVEVEKEGATQWRLCHELVGPEVCVWGLAPGETY | 4592 |
| Qy | 3887 | ----- | 3886 | Db | 4593 | RFRVAAGVPVGAEPVHLPTQVRLABPPKPPQPSAPESROVAAGEDVSELEVVVAEAG | 3176 |
| Db | 2161 | QIRGLAVEDTGEYLCVCGOERTSATLTVRALPARFIDNMNTNOEAREGATATLHCBSKVA | 2220 | Qy | 3177 | RFRVAAGVPVGAEPVHLPTQVRLABPPKPPQPSAPESROVAAGEDVSELEVVVAEAG | 4652 |
| Qy | 3887 | ----- | 3886 | Db | 4653 | EVTHKGMERIPGGRFEVVSQGRQOMLVIKGFTABDQGEYHCGLAGGSCIPAAATFOVA | 3235 |
| Db | 2221 | PVWVRKGPETLRDGRHSRLQDGRCELOIHLGLSVADAGEYSCVCGOERTSATLTIRALP | 2280 | Qy | 3236 | EVTHKGMERIPGGRFEVVSQGRQOMLVIKGFTABDQGEYHCGLAGGSCIPAAATFOVA | 4712 |
| Qy | 3887 | ----- | 3886 | Db | 4713 | LSPASVDEAPQPSLPPEAAQEGDLHLWEALARKRMSREPTLDSISELPEEDGRQORLP | 3295 |
| Db | 2281 | AKFTKGLRNEATEGATAMLOCELSKVAVPVMRKGPELTRDGRYNLQDGRCELOIHLG | 2340 | Qy | 3296 | LSPASVDEAPQPSLPPEAAQEGDLHLWEALARKRMSREPTLDSISELPEEDGRQORLP | 4772 |
| Qy | 3887 | ----- | 3886 | Db | 4773 | QEAEEVAPDLSEGYSTADELARTGDADLSHTSSDDESAGTSLVTLKXKAGRPQTSPLA | 3355 |
| Db | 2341 | LSVADTGEYSCVCGOERTSATLTVRAPQPVFRPEPLQSLQABEGSTATLQCSLSEPTATV | 2400 | Qy | 3356 | QEAEEVAPDLSEGYSTADELARTGDADLSHTSSDDESAGTSLVTLKXKAGRPQTSPLA | 4832 |
| Qy | 3922 | WSKGGIQLQANGRRPRLOQCTAELVLQDLOREDTGEYTCGSOATSATLTITVAAPVRF | 3981 | Db | 4833 | SKVGPAP----- | 3415 |
| Db | 2401 | WSKGGIQLQANGRRPRLOQCTAELVLQDLOREDTGEYTCGSOATSATLTITVAAPVRF | 2460 | Qy | 3416 | SKVSPNACKERFPTPRAGSLLGFVGADPAPFGSERSACTRRCAAPPRESLKRREP | 4852 |
| Qy | 3982 | LRELQHOEVEGTAHLCCELSRAGASVWVRKGSLOLPFCAYQMVQDGAALLVRGVE | 4041 | Db | 4853 | AAVRPPLGLDLS TKDLG----- | 3474 |
| Db | 2461 | LRELQHOEVEGTAHLCCELSRAGASVWVRKGSLOLPFCAYQMVQDGAALLVRGVE | 2520 | Qy | 3475 | ASCLP--GAMEAVELARKLOEBEATCICLDYFDTPVMTTCGHNFRCACIQLSWEKARGKX | 4884 |
| Qy | 4042 | QEDAGDYTCCTGHTOSMASLSVRVRPKFKTLQSLQETGTDIARLCCQLSDAESGAVVQ | 4101 | Db | 4885 | GYKVRK-----EMKQOEGPMFHTFGDTAQVGDALRLCEVVASKADVBARWLKDG | 3532 |
| Db | 2521 | QEDAGDYTCCTGHTOSMASLSVRVRPKFKTLQSLQETGTDIARLCCQLSDAESGAVVQ | 4161 | Qy | 3533 | GRKRKGSFPCECREMSQORNLN----- | 4395 |
| Qy | 4102 | WLKEGVHLAGPKYEMRSQATRELLIHOLEAKDTGEYACVTGGOKTAAASURVTEPEVTI | 4161 | Db | 4396 | VELTDGHHHIDQLGDGTCSLIAGLDADAGCVTCQVSNK----- | 3568 |
| Db | 2550 | -----CGPQVRDAAQAGATRELLIHOLEAKDTGEYACVTGGOKTAAASURVTEPEVTI | 2600 | Qy | 4397 | PGLOKQD----- | 4991 |
| Qy | 4162 | VRGLVDAEVTADVEDFEFSCEVRAGATGVOMCLOGLPLOSNEVTEVAVRDGRIHTLRKLG | 4221 | Db | 3569 | -----COH----- | 3604 |
| Db | 2601 | VRGLVDAEVTADVEDFEFSCEVRAGATGVOMCLOGLPLOSNEVTEVAVRDGRIHTLRKLG | 2660 | Qy | 4992 | ESEAESGGGLDDAFRAAARRHLRFRKSPAEVSDDEELFLSADEG----- | 5038 |
| Qy | 4222 | VTPEAGTVSFHLGNHASSAQLTVRAPEVTILEPLQDVQLR | 4281 | Db | 3605 | SRE----- | 3643 |
| Db | | | | Qy | 5039 | -----PABPEPADWQ--TYREDEHFICIRFEAL----- | 5079 |
| | | | | Db | 3644 | LOAREEOSLAEBWQGVKERRERIVLEPEKVNLYLVEEQRLLOALETEEBETASLRRESV | 3703 |

QY 5080 ATL---GIGVEIKLVE-----QGRPR-----VEMCISK-- 5104
Db 3704 ACLDRGHSLELLQLLEERSTQGLQMLQDKMKEPLSRAALLVLIHGMNVLVEFPVSLP 3763
QY 5105 SPLYLIATKAHTQLGPGTTPDEPCTPLPISPP-PRPS--TEDVVPDATSAFYPLLVE 5131
Db 3764 SPLYLIATKAHTQLGPGTTPDEPCTPLPISPP-PRPS--TEDVVPDATSAFYPLLVE 3820
QY 5132 LQNEVEQDGP-----VSPDCVVTQOPNPSVRWFKDGKLLLEDHYMINEQOQ 5180
Db 3821 SRQRYLSSPGSGFCSDKRFVAYPCAV-QGTA-----FSSGR-----HYNEVGMNIT 3868
QY 5181 GHQLIITAVPADMG-----VYRCIAENSMGVSSTKAELRVDLTSTDYDTAADATESSYF 5236
Db 3869 GDALWALGVCRDNVSRKDRVPKC-PENGFVW-----VQLSK-----GTXYLSTP 3911
QY 5237 SAQGYLSRE---QEGTESTTDEGQLPOVVEELRDQLVAPGTRLAKTQLKVKGYAPAPRLY 5293
Db 3912 SALTPVWMLPEPSHMGIFLDFEAG-----EVSFYSVSDGSHLHTYSQAT--PPGGLQP 3962
QY 5294 WFKDQPLTASAHIRMTGKILHTLEIISVTRSDSGOYAAVYSNAGAAVSSARLLVRGP 5353
Db 3963 FFCLGAP-----KSGQWISVTVYAGV-KDLATRTGAVVTPALGA----- 4001
QY 5354 DEPEKPASDVHEQLVPPMLERTPKKKKGSSITFSVKVEGRPVPTVHMLREAPRGV 5413
Db 4002 -----YAPGATETQS-----PAP-----WSPRAPE--- 4021
QY 5414 LWIGPDTGYTVASSAQHSLVLLDVGROHGTCTCIASNAAGQALCSASLHVSLPKVE 5473
Db 4022 -----PEHFG--VPSLAPRSA-----RACAAAPGYGSPRAA 4051
QY 5474 BOEKVEKALISTFLOGTQTAISAOGLTASADIGQORKEEPLAAKEALGHLSLAEVGT 5533
Db 4052 EAARRRPADSTAFLP-SVRAMAA-----PDLSTNLQEAATCA----- 4087
QY 5534 EFLQKLSQITEMYSAKITQAKLVQPGDSDDESKTPSAS--PRHGRSPSSI----- 5585
Db 4088 ICLDYFTDPVMDCGHNFRCIRCGQPEARTRAPASCPRGTCGPTARLLRWPRW 4147
QY 5586 QESSESESDGARGEIIDIYVVTADYPLGAE-QDAITL-----REGQVVEVLDAHPL 5638
Db 4148 RGACTRRRRRR-----VPAHREPLAFCDELLCAACERSGEH----- 4189
QY 5639 RWL-----VRTKPTKSPSRQ-----WSPAYLDRRLK-LSPE----- 5671
Db 4190 -WAHRVGRCTRTPKTSRPLEAGTMAANETLLSGAKLEKSLHLRKQWQDALLFQAQDET 4248
QY 5672 ---WGAAEAEFPGEAEVSEYKARLSSVIOELLSAQAFVELOFLOSHHL-----QHLE 5724
Db 4249 CVLWQADGGEQORQNVLEFE---RLRLLAEGTAAAEAGBELKQSAHLAELE 4305
QY 5725 RCPHVPIAVAGQKA 5738
Db 4306 R--PLPAACAGAAA 4317

RESULT 5

ABP70085

ID ABP70085 standard; protein; 4675 AA.

XX AC ABP70085;

XX DT

XX 27-JAN-2003 (first entry)

XX Human NOV13b.

XX Human; anti-HIV; cytostatic; antidiabetic; antiasthmatic; cachexia; AIDS;

KW antiinflammatory; cardiac; haemostatic; neuroprotective; anorectic;

KW notropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;

KW antifertility; cerebroprotective; gene therapy; NOVX; NOV; fertility;

KW metabolic disorder; diabetes; obesity; infectious disease; anorexia;

KW neurodegenerative disease; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; cardiovascular disorder;
KW bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;
KW metabolic syndrome X; wasting disorder; cell differentiation;
KW cell proliferation; haematopoiesis; wound healing; angiogenesis.
OS Homo sapiens.
XX WO200272771-A2.
XX 19-SEP-2002.
XX 08-MAR-2002; 2002WO-US007288.
XX 08-MAR-2001; 2001US-0274101P.
XX 08-MAR-2001; 2001US-0274194P.
XX 08-MAR-2001; 2001US-0274281P.
XX 08-MAR-2001; 2001US-0274322P.
XX 09-MAR-2001; 2001US-0274849P.
XX 12-MAR-2001; 2001US-0275235P.
XX 13-MAR-2001; 2001US-0275578P.
XX 13-MAR-2001; 2001US-0275797P.
XX 13-MAR-2001; 2001US-0275601P.
XX 14-MAR-2001; 2001US-0276000P.
XX 16-MAR-2001; 2001US-0276767P.
XX 20-MAR-2001; 2001US-0276994P.
XX 20-MAR-2001; 2001US-0277239P.
XX 20-MAR-2001; 2001US-0277321P.
XX 20-MAR-2001; 2001US-0277327P.
XX 20-MAR-2001; 2001US-0277388P.
XX 21-MAR-2001; 2001US-0277791P.
XX 22-MAR-2001; 2001US-0277833P.
XX 23-MAR-2001; 2001US-0278152P.
XX 26-MAR-2001; 2001US-0278894P.
XX 27-MAR-2001; 2001US-0278999P.
XX 27-MAR-2001; 2001US-0279036P.
XX 28-MAR-2001; 2001US-0279344P.
XX 30-MAR-2001; 2001US-0279995P.
XX 30-MAR-2001; 2001US-0280233P.
XX 02-APR-2001; 2001US-0280802P.
XX 02-APR-2001; 2001US-0280822P.
XX 02-APR-2001; 2001US-0280900P.
XX 04-APR-2001; 2001US-0281194P.
XX 13-APR-2001; 2001US-0283675P.
XX 30-APR-2001; 2001US-0287424P.
XX 02-MAY-2001; 2001US-0288066P.
XX 03-MAY-2001; 2001US-0288342P.
XX 15-MAY-2001; 2001US-0288528P.
XX 16-MAY-2001; 2001US-0291039P.
XX 16-MAY-2001; 2001US-0291240P.
XX 30-MAY-2001; 2001US-0294485P.
XX 31-MAY-2001; 2001US-0294899P.
XX 18-JUN-2001; 2001US-0299027P.
XX 19-JUN-2001; 2001US-0299303P.
XX 19-JUN-2001; 2001US-0299310P.
XX 10-JUL-2001; 2001US-0304334P.
XX 31-JUL-2001; 2001US-0309198P.
XX 16-AUG-2001; 2001US-0312903P.
XX 10-SEP-2001; 2001US-0318462P.
XX 12-SEP-2001; 2001US-0318770P.
XX 27-SEP-2001; 2001US-0325430P.
XX 27-SEP-2001; 2001US-0325681P.
XX 18-OCT-2001; 2001US-0330380P.
XX 31-OCT-2001; 2001US-0335301P.
XX 14-NOV-2001; 2001US-0332172P.
XX 14-NOV-2001; 2001US-0332271P.
XX 14-NOV-2001; 2001US-0332272P.
XX 14-NOV-2001; 2001US-0333184P.
XX 21-NOV-2001; 2001US-0333272P.
XX 21-NOV-2001; 2001US-0332094P.
XX 03-DEC-2001; 2001US-0337426P.

PR 03-DEC-2001; 2001US-0338092P.
 PR 04-DEC-2001; 2001US-0337185P.
 PR 03-JAN-2002; 2002US-0345705P.
 PR 08-MAR-2002; 2002US-00093463.
 XX (CURA-) CURAGEN CORP.
 PA Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;
 PI Boldog FL, Li L, Zerhusen BD, Tchernev VT, Gangolli EA, Vernet CAM;
 PI Pena CEA, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;
 PI Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;
 PI Taupier RJ, Padigaru M, Shenoy SG, Kekuda R, Gusev VI, Pochart PF;
 PI Zhong M;
 XX WPI; 2002-732824/79.
 DR N-PSDB; ABV99363.
 XX New NOVX polypeptides and polynucleotides, useful for preventing,
 PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,
 PT Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic
 PT disorders, and asthma.
 XX Claim 1; Page 142-143; 619pp; English.
 XX The present invention relates to new isolated proteins (NOVX) and their
 CC coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is
 CC any number from 1 to 48. The NOVX proteins and coding sequences are
 CC useful in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, preferably a NOVX-associated disorder.
 CC The NOVX coding sequences and proteins are useful for treating, diabetes,
 CC preventing or diagnosing diseases such as metabolic disorders, diabetes,
 CC obesity, infectious disease, anorexia, cancer-associated cachexia,
 CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
 CC disease, immune disorders, hematopoietic disorders, cardiovascular
 CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
 CC disturbances associated with obesity, metabolic syndrome X or wasting
 CC disorders associated with chronic diseases or various cancers. The NOVX
 CC coding sequences and proteins may also be used as targets for the
 CC identification of small molecules that modulate or inhibit e.g.
 CC neurogenesis, cell differentiation, cell proliferation, haematopoiesis,
 CC wound healing and angiogenesis, in gene therapy, in generation of
 CC antibodies that bind immunospecifically to NOVX substances for use in
 CC therapeutic or diagnostic methods
 XX Sequence 4675 AA;
 SQ
 Query Match 25.5%; Score 10506; DB 5; Length 4675;
 Best Local Similarity 56.1%; Pred. NO. 0;
 Matches 2252; Conservative 45; Mismatches 129; Indels 1586; Gaps 35;
 QY 2595 MPLYNDSPHEISHKGRRTLVLSKIQRADAGIVRASSLKVSTARSARLEVRKPVVFLKALD 2654
 DB 1 MPLYNDSPHEISHKGRRTLVLSKIQRADAGIVRASSLKVSTARSARLEVRKPVVFLKALD 60
 QY 2655 DLSAEERTGLAQCEVSPFAHVVRKDGVLGSPDKYDFLHTAGTGLVVDVSPEDAG 2714
 DB 61 DLSAEERTGLAQCEVSPFAHVVRKDGVLGSPDKYDFLHTAGTGLVVDVSPEDAG 120
 QY 2715 LYTCHVGSEETRARVRVHDHLVGTIKTKTMEVLEGESCSPECVLSHSASDPANVTGG 2774
 DB 121 LYTCHVGSEETRARVRVHDHLVGTIKTKTMEVLEGESCSPECVLSHSASDPANVTGG 180
 QY 2775 KTVGSSRFQATROGRKYLIVVREAPSDAGEVVFVRGLTSKASLIVRERPAAIIXPLE 2834
 DB 181 KTVGSSRFQATROGRKYLIVVREAPSDAGEVVFVRGLTSKASLIVRERPAAIIXPLE 240
 QY 2835 DQWAPGEDVELRCELSRAGTPVHVKDKAIRKSQKYDVVCEGTMMAMLVIRGASLKDAG 2894
 DB 241 DQWAPGEDVELRCELSRAGTPVHVKDKAIRKSQKYDVVCEGTMMAMLVIRGASLKDAG 300
 QY 2895 EYTCVEASKSTASLHVEKANCFTTELTNLQVEKGTAVFTCKTEHPAAVTWVRKGLLE 2954
 DB 301 EYTCVEASKSTASLHVEKANCFTTELTNLQVEKGTAVFTCKTEHPAAVTWVRKGLLE 360

QY 2955 LRASGKHQPSQEGTLRLTISALEKADSDTYTCDIGQAQSRQAQLLVQGRVHIIEDLEV 3014
 DB 361 LRASGKHQPSQEGTLRLTISALEKADSDTYTCDIGQAQSRQAQLLVQGRVHIIEDLEV 420
 QY 3015 DVQEGSSATFRCRISPANYEPVHWFELDKTPLHANELNEIDAQPGGYHVLTLROLAKDSG 3074
 DB 421 DVQEGSSATFRCRISPANYEPVHWFELDKTPLHANELNEIDAQPGGYHVLTLROLAKDSG 480
 QY 3075 TIYFAGDORASAAALRVTEKPSVFSRELTLDATITEGEDTLVCESTSTCDIPMCWTKDGT 3134
 DB 481 TIYFAGDORASAAALRVTEKPSVFSRELTLDATITEGEDTLVCESTSTCDIPMCWTKDGT 540
 QY 3135 LRGSARCOLSHGHEHRAQLLITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFQALKDL 3194
 DB 541 LRGSARCOLSHGHEHRAQLLITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFQALKDL 600
 QY 3195 EVLEGGATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVNRLPDQSGRY 3254
 DB 601 EVLEGGATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVNRLPDQSGRY 660
 QY 3255 SCSFGDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAAPVEMRKGSSETLRD 3314
 DB 661 SCSFGDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAAPVEMRKGSSETLRD 720
 QY 3315 GDRYCLRQDGMANCELOIRGLAMVDAAYSCVCGEERTSASLTIRPMPAHFICGLRHQESI 3374
 DB 721 GDRYCLRQDGMANCELOIRGLAMVDAAYSCVCGEERTSASLTIRPMPAHFICGLRHQESI 780
 QY 3375 EGATATLRCELSKTAAPVEMRKGSSETLRDQDGRHSRLRQDGAVALCELOICGLAVADAGEYSCVC 3434
 DB 781 EGATATLRCELSKTAAPVEMRKGSSETLRDQDGRHSRLRQDGAVALCELOICGLAVADAGEYSCVC 840
 QY 3435 GEERTSATLTVTALPAKFTFEGLRNBEAVEGATAMLWCELSKVAPVEMRKGPENLDGDRY 3494
 DB 841 GEERTSATLTVTALPAKFTFEGLRNBEAVEGATAMLWCELSKVAPVEMRKGPENLDGDRY 900
 QY 3495 ILRQEGTRCELOICGLAVADAGEYLCVCGQERTSATLTIRALPARFIEDVKQNEAREGAT 3554
 DB 901 ILRQEGTRCELOICGLAVADAGEYLCVCGQERTSATLTIRALPARFIEDVKQNEAREGAT 960
 QY 3555 AVLQCELSNAAAPVEMRKGSSETLRDGDYSLRDQGTKELOIRGLAMADTGEYSCVCGQBR 3614
 DB 961 AVLQCELSNAAAPVEMRKGSSETLRDGDYSLRDQGTKELOIRGLAMADTGEYSCVCGQBR 1020
 QY 3615 TSAMLTVRALPIKFTFEGLRNBEATEGATAVLRCELSKMAAPVEMWKGHETLRDGRHSRLQ 3674
 DB 1021 TSAMLTVRALPIKFTFEGLRNBEATEGATAVLRCELSKMAAPVEMWKGHETLRDGRHSRLQ 1080
 QY 3675 DGARCELOIRGLVAEDAGEYLCVCGKERTSAMLTVRAMPKSFIEGLRNEEATEGDTATLW 3734
 DB 1081 DGARCELOIRGLVAEDAGEYLCVCGKERTSAMLTVRAMPKSFIEGLRNEEATEGDTATLW 1140
 QY 3735 CELSKAAPVEMRKGHETLRDGRHSRLRQDGRSCELOIRGLAVVDAGEYSCVCGQERTSAT 3794
 DB 1141 CELSKAAPVEMRKGHETLRDGRHSRLRQDGRSCELOIRGLAVVDAGEYSCVCGQERTSAT 1200
 QY 3795 LTVRALPARFIEDVKQNEAREGATAVLQCELSKAAAPVEMRKGSSETLRGGDRYSLRDGTR 3854
 DB 1201 LTVRALPARFIEDVKQNEAREGATAVLQCELSKAAAPVEMRKGSSETLRGGDRYSLRDGTR 1260
 QY 3855 CELQIHGLSVADTGEYSCVCGQERTSATLTVR ----- 3886
 DB 1261 CELQIHGLSVADTGEYSCVCGQERTSATLTVRALPARFTQDLKTKASEGATATLQCELS 1320
 QY 3887 ----- 3886
 DB 1321 KVAPVEMWKGPETLRDGRYSLKQDGTCELOIHLDSVADAGEYSCMCGQERTSATLTVR 1380
 QY 3887 ----- 3886
 DB 1381 ALPARFTEGLRNEEAMEGATATLQCELSKAAAPVEMRKGLEALRDGDKYSLRDGAVCELO 1440

QY 3887 ----- 3886
Db 1441 IHGLAMADNGVYSCVCGQERTSATLTVRALPARFIEDMRNOKATEGATVTLQCKLRKAAP 1500
QY 3887 ----- 3886
Db 1501 VEWKGPNTLKDGRYSLKQDGTSCELQIRGLVIADAGEYSCICEQERTSATLTVRALPA 1560
QY 3887 ----- 3886
Db 1561 RPIEDVRNHEATEGATAVLQCELSKAAPEVWRKGSSETLRDGRYSLRQDGTGRCELQIRGL 1620
QY 3887 ----- 3886
Db 1621 AVEDTGEYLCVCGQERTSATLTVRALPARFIDNMTNQBAREGATATLHCELSKVAPVEWR 1680
QY 3887 ----- 3886
Db 1681 KGPETLRDGRHSRLRQDGRSCELQIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFIE 1740
QY 3887 ----- 3886
Db 1741 DVYQNEAREGATAVLQCELSKAAPEVWRKGSSETLRDGRYSLRQDGTGRCELQIRGLSVAD 1800
QY 3887 ----- 3886
Db 1801 TGEYSCVCGQERTSATLTVRALPARFTQDLTKTEASEGATATLQCELSKVAPVEWKGP 1860
QY 3887 ----- 3886
Db 1861 TLRDGRYSLKQDGTGRCELQIHDLSVADAGEYSCMCGQERTSATLTVRDCHTLVMPHY 1920
QY 3887 ----- 3886
Db 1921 FQLPGLKEPBETLIYQIPSPVILFTEGURNEAMEGATATLQCELSKAAPEVWRKGLE 1980
QY 3887 ----- 3886
Db 1981 ALRDGKYSLRQDGAVCBLQIHGLAMADNGVYSSLPARFIEDMRNOKATEGATVTLQCKL 2040
QY 3887 ----- 3886
Db 2041 RKAAPVEWRKGPNTLKDGRYSLKQDGTSCELQIRGLVIADAGEYSCICEQERTSATLTV 2100
QY 3887 ----- 3886
Db 2101 RALPARFIEDVRNHEATEGATAVLQCELSKAAPEVWRKGSSETLRDGRYSLRQDGTGRCEL 2160
QY 3887 ----- 3886
Db 2161 QIRGLAVEDTGEYLCVCGQERTSATLTVRALPARFIDNMTNQBAREGATATLHCELSKVA 2220
QY 3887 ----- 3886
Db 2221 PVWRKGPETLRDGRHSRLRQDGTGRCELQIRGLSVADAGEYSCVCGQERTSATLTVRALP 2280
QY 3887 ----- 3886
Db 2281 AKFTKGLRNEBATEGATAMLQCELSKVAPVEWRKGPETLRDGRYNLRQDGTGRCELQIHG 2340
QY 3887 ----- 3886
Db 2341 LSVADTGEYSCVCGQERTSATLTVKAPQVPFRPLOSQAEEGSTATLQCELSEPTATVV 2400
QY 3922 WSKGGLQLANGREPRLOQCTAELVLQDLQREDTGEYTCGQSQATSATLTVTAAPVRF 3981
Db 2401 WSKGGLQLANGREPRLOQCTAELVLQDLQREDTGEYTCGQSQATSATLTVTAAPVRF 2460
QY 3982 LRELQHOEVDGGTALHCELSRAGASVEWRKGSLOLFPCKAKYQMYQDGAABELLVRGVE 4041
Db 2461 LRELQHOEVDGGTALHCELSRAGASVEWRKGSLOLFPCKAKYQMYQDGAABELLVRGVE 2520
QY 4042 QEDAGDYTCDTGHTQTSVASLSVRPFRPKFKTRILQSLQEQETGDIARLCCQLSDAESGAVVQ 4101

Db 2521 QEDAGDYTCDTGHTQTSVASLSVRGGR-----GAA-- 2549
QY 4102 WLKEGVELHAGPKYEMRSQGATRELLIHOLEAKOTGEYACVTGGOKTAASLRVTEPEVTI 4161
Db 2550 -----CGPOVRDAAGATRELLIHOLEAKOTGEYACVTGGOKTAASLRVTEPEVTI 2600
QY 4162 VRGLVDAEYTADEDEFFSCVSRAGATGVQWCLQGLPLQSNVEVTVAVRDRGIHTLRLKG 4221
Db 2601 VRGLVDAEYTADEDEFFSCVSRAGATGVQWCLQGLPLQSNVEVTVAVRDRGIHTLRLKG 2660
QY 4222 VTPEDAGTVSFHLGHASQAQLTVRAPEVTILEPQDVOLSEGQDASFOCRLSRASGOEA 4281
Db 2661 VTPEDAGTVSFHLGHASQAQLTVRAPEVTILEPQDVOLSEGQDASFOCRLSRASGOEA 2701
QY 4282 RWALGCVPLQANEMNDITVEQGTLLHLTLHKVTLEDACTVSHVGTCSSEALQKVT--- 4337
Db 2702 -----GVPLQANEMNDITVEQGTLLHLTLHKVTLEDACTVSHVGTCSSEALQKVTAEVP 2756
QY 4338 ----- 4337
Db 2757 CLVRGLQNVDPAGEVATFSCEDGQSAIAVRDGIHFHLSMLSLGLGVADSGTVIFRAGPLV 2816
QY 4338 ----- 4337
Db 2817 STAKLLIKDPVVVVVSAMQDLAVBEGGSABELLCQYSRPVQATWMDEREVTDGHRVIE 2876
QY 4338 -----AKNTVVRGLENVEALEGEALFE 4360
Db 2877 QDMNVARLTFRPAIPCDSGIYSCBAAGTRVVALLQVQAKNTVVRGLENVEALEGEALFE 2936
QY 4361 COLSQPEVAANTWLLDDPEVTSNAEAVVPENGIRHLILLIKNLRPODSCHVTFLAGDW 4420
Db 2937 COLSQPEVAANTWLLDDPEVTSNAEAVVPENGIRHLILLIKNLRPODSCHVTFLAGDW 2996
QY 4421 TSAFLTVR-----GWRLEILEPLKNAAVRAGAQAARFTCTLSEAVPVGEASVINGAA 4472
Db 2997 TSAFLTVRGDCAVLQVQWRLEILEPLKNAAVRAGAQAARFTCTLSEAVPVGEASVINGAA 3056
QY 4473 VOPDDSDWTVTADGSHQALLRSQPHHAGEVTFACRDNAVASARLTVLGLPDPPEDAEV 4532
Db 3057 VOPDDSDWTVTADGSHQALLRSQPHHAGEVTFACRDNAVASARLTVLGLPDPPEDAEV 3116
QY 4533 AHSHTVTLNAAWPMSPMSDGGGLCGYRVEVEKAGATQWRLCHELVGPECVVDGLAPGETY 4592
Db 3117 ARSHTVTLNAAWPMSPMSDGGGLCGYRVEVEKAGATQWRLCHELVGPECVVDGLAPGETY 3176
QY 4593 RFRVAAGPVGAGBPVHLPOTVRLAEPKPVPPQPSAPESRQVAAAGEDVSLSEVVAEAG 4652
Db 3177 RFRVAAGPVGAGBPVHLPOTVRLAEPKPVPPQPSAPESRQVAAAGEDVSLSEVVAEAG 3235
QY 4653 EVIWHKGMERTQPGGRPEVVSQGRQOMLVIKGFTAEQDQGEYHCCGLOAGSICPAAATFOVA 4712
Db 3236 EVIWHKGMERTQPGGRPEVVSQGRQOMLVIKGFTAEQDQGEYHCCGLOAGSICPAAATFOVA 3295
QY 4713 LSPASVDEAPQPSLPPEAAQEGDLHLWEALARKRMSREPTLDSISELPEEDGRSRLP 4772
Db 3296 LSPASVDEAPQPSLPPEAAQEGDLHLWEALARKRMSREPTLDSISELPEEDGRSRLP 3355
QY 4773 QEAEVAPDLSEGYSTADELARTGDADLSHTSSDDESRAGTSPSLVTVLKKAGRGTSPLA 4832
Db 3356 QEAEVAPDLSEGYSTADELARTGDADLSHTSSDDESRAGTSPSLVTVLKKAGRGTSPLA 3415
QY 4833 SKVGAP-----AAPSVKPQQOQEPL 4852
Db 3416 SKVSPNNLACKERFPTFRAGSLILGVGADPAPFGSERSARCTRCAAPPPRESLKEP- 3474
QY 4853 AAVRPPGLDLSKDLG-----DPSMD-----KAAVKIQ-----AAFK 4884
Db 3475 ASCIPL--GAMEAVELARKLOEAEATCS:CLDYFTDPMVTTGHNFCRACIQLSWEKARGKK 3532
QY 4885 GYKVRK-----ENKQOEGPMFSTFGDTAQVGDALRLCEVVASKADVVRWMLKDG 4935

Db 3533 GRKRKGSFPCECKREMSQPNLLPN-----RLLTQVAEMA----- 3568
 QY 4936 VELTDGRHHHDQLGDCGTCSLLIAGLORADAGCYTCQVSNK-----FCQVTHSACVVVSGS 4991
 Db 3569 -----QQH-----PGLQKQD-----LCQEHHEPLKLFCKQDQSPICVVCVRE 3604
 QY 4992 ESEAESSGGELDDAFRAARRHLRLFRKTSAPAEVSDDEELFSADEG----- 5038
 Db 3605 SRE-----HRLHVL-----PAEEAVQGYKLEEDMEVYLREQITRTGN 3643
 QY 5039 --PAPEPAPDWO--TYREDEFICIRFEAL-----TEARQAVTRFQEMF 5079
 Db 3644 LQAREEQSLAEQGVKXERRRIVLEFEKNNLYLVEEQRLLQALETEEBEASRLRESV 3703
 QY 5080 ATL--GIGVRIKLVE-----QGPRR-----VENCISK-- 5104
 Db 3704 ACLDRQGHSLLELLQLLEERSTQGLQMLQDMKEPLSRAALLVLIHGNNLVEFPVSLP 3763
 QY 5105 -----ETPAVVPPEPLPSLTSDAAP-----VPLTE 5131
 Db 3764 SPLYLIAKATQLGPGTTFDPECPTPLPISPP-PRPS--TEDVVPDATSAYPEYLLLYE 3820
 QY 5132 LQNEQVQDGYP-----VSFDQVVTGQPMPSVRWPKDGKXLEEDDHY 5172
 Db 3821 SRQRYLGSSPEGSGFCSDRFVAYPCAV-QGTA-----FSSGR-----HY 3860

RESULT 6
 ID ABP58227 standard; protein; 2328 AA.
 XX ABP58227;
 AC ABP58227;
 DT 31-MAR-2003 (first entry)
 XX
 XX Human cell adhesion and extracellular matrix protein 4.
 XX
 KW Cell adhesion and extracellular matrix protein 4; CADECM-4; human;
 KW anti-Hiv; virucide; antiallergic; antiinflammatory; antianaemic;
 KW antiparkinsonian; nootropic; anticonvulsant; antiinfertility;
 KW antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;
 KW cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;
 KW antigout; thyromimetic; neuroprotective; osteoprotic; antiarthritis;
 KW antiparasitic; antihelminthic; antipsoiatic; uropathic; ophthalmological;
 KW antirheumatic; haemostatic; antibacterial; protozoacide; fungicide;
 KW gynaecological; neurexin; gene therapy.
 XX
 XX Homo sapiens.
 OS
 XX WO200288322-A2.
 PN
 XX
 XX 07-NOV-2002.
 PD
 XX
 XX 01-MAY-2002; 2002WO-US013874.
 PF
 XX
 XX 02-MAY-2001; 2001US-0289290P.
 PR
 XX 21-MAY-2001; 2001US-0292468P.
 PR
 XX 15-JUN-2001; 2001US-0298616P.
 PR
 XX 28-JUN-2001; 2001US-0301672P.
 PR
 XX 04-JAN-2002; 2002US-0345008P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX
 XX Yue H, Lee EA, Duggan BM, Thangavelu K, Honchell CD, Ding L;
 PI Hillman JL, Baughn MR, Kallick DA, Lee S, Warren BA, Xu Y, Tran UK;
 PI Lal PG, Thornton M, Hafalia Aja, Yao MG, Nguyen DB, Gandhi AR;
 PI Khan FA, Walhia NK, Griffin JA, Chinn AM, Elliott VS, Ramkumar J;
 PI Arvizu CS, Forsythe IJ;
 XX
 XX WPI; 2003-167112/16.
 DR
 XX N-PSDB; AB224581.
 DR
 XX New human cell adhesion and extracellular matrix proteins, useful for

PT diagnosing, treating or preventing autoimmune or inflammatory disorder
 PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,
 PT cancer or hepatitis.
 XX
 XX Claim 1; Page 144-150; 178pp; English.
 XX
 XX The present sequence is the protein sequence of Incyte polypeptide
 7326129CD1 denoted human cell adhesion and extracellular matrix protein 4
 CC (CADECM-4). The protein is encoded by a clone isolated from a male muscle
 CC cDNA library. Homology searches indicate it to be a titin muscle protein.
 CC The invention provides CADECM-1 to -11 polypeptides (see ABP58224-34) and
 CC polynucleotides (see AB224578-88), expression vectors, host cells,
 CC antibodies, agonists and antagonists. These are useful for diagnosing,
 CC treating or preventing disorders associated with aberrant expression of
 CC CADECM, particularly cell proliferative disorders (e.g. arteriosclerosis,
 CC atherosclerosis, cirrhosis, hepatitis, paroxysmal nocturnal
 CC haemoglobinuria, polycythaemia vera, psoriasis, primary
 CC thrombocytopenia or cancer), developmental disorders (e.g. renal
 CC tubular acidosis, anemia or mental retardation), neurological disorders
 CC (e.g. Alzheimer's disease, Parkinson's disease or epilepsy), reproductive
 CC disorders (e.g. infertility or a disruption in the menstrual cycle), or
 CC autoimmune/inflammatory disorders (e.g. AIDS, allergy, asthma, autoimmune
 CC thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus,
 CC glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease,
 CC Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
 CC osteoarthritis, osteoporosis, pancreatitis, Reiter's syndrome, rheumatoid
 CC arthritis, Sjogren's syndrome, uveitis), or viral, bacterial, fungal,
 CC parasitic, protozoal or helminthic infections
 XX
 XX Sequence 2328 AA;
 QY
 Query Match 23.8%; Score 9834.5; DB 6; Length 2328;
 Best Local Similarity 84.6%; Pred. No. 0;
 Matches 1941; Conservative 76; Mismatches 213; Indels 65; Gaps 8;
 QY 2096 MEVQLSHADVDGSGWTRDGLRFQOGPTCHLAVRGPMTLLSLRPPDSGLMVFKAEGVHT 2155
 Db 1 MEVQLSHADVDGSGWTRDGLRLQGPPTCHLAVRGPMTLLSLRPPDSGLMVFKAEGVHT 60
 QY 2156 SARLVVTELPVPSRPLQDVVTEKVKVLELSPNVVVRWKDGVLRAGKTWAIAA 2215
 Db 61 SARLVVTELPVPSRPLQDVVTEKVKVLELSPNVVVRWKDGVLRAGKTWAIAA 120
 QY 2216 QGACRSLLTYRCFAQGVVVCDAHDAQSSASVKVQ----- 2251
 Db 121 QGACRSLLTYRCFAQGVVVCDAHDAQSSASVKVQGRNIQIVRPLEDEVVEVKDGATFS 180
 QY 2252 -----GRTYLLIYRVLAEADAGEIQFVAENAESRA 2281
 Db 181 CEVSHDEVPGQWFEGSKLRPTDNRIRQEGRTYLLIYRVLAEADAGEIQFVAENAESRA 240
 QY 2282 QLRVKELPVTLPRLDKIAVEKHGVLQCOVSRAQVRFKGSQELQPGPKYELVSDG 2341
 Db 241 QLRVKELPVTLPRLDKIAVEKHGVLQCOVSRAQVRFKGSQELQPGPKYELVSDG 300
 QY 2342 LYRKLIISDVHAEDDTYTCADGDVKTSAQFFVEEQSIITVIRGLQDVVMEPAPAFCE 2401
 Db 301 LYRKLIISDVHAEDDTYTCADGDVKTSAQFFVEEQSIITVIRGLQDVVMEPAPAFCE 360
 QY 2402 TSIPSVRPPKWLIGKTVLQAGGNVGLFOEGTVHLMRLRTCSMTGCVHFTVCKSSSAR 2461
 Db 361 TSIPSVRPPKWLIGKTVLQAGGNVGLFOEGTVHLMRLRTCSMTGCVHFTVCKSSSAR 420
 QY 2462 LVVSDIPVVLTRPLEPKTGRELQSVVLSGDFRPAKAVQWKDDTPLSPSEKFKMSLEGQ 2521
 Db 421 LVVSDIPVVLTRPLEPKTGRELQSVVLSGDFRPAKAVQWKDDTPLSPSEKFKMSLEGQ 480
 QY 2522 MAELRILRLMPADAGVYRCQAGSAHSSTVTVAREVTVTGPLQDAEAEEGHNASFCSEL 2581
 Db 481 MAELRILRLMPADAGVYRCQAGSAHSSTVTVAREVTVTGPLQDAEAEEGHNASFCSEL 540
 QY 2582 SHDEEVEWELNGMPLYNDSFHEISHKGRHHTLVLSIORADAGIVRASLSKYSTSARLE 2641

Db 541 SHDEEVEWNGMPLYNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTSRLE 600
 Qy 2642 VRVKEVFLKALDDLSABERGTLALQCEVSDPEAHVVRKDGVOQLGPSDKYDFLHTAGTR 2701
 Db 601 VRVKEVFLKALDDLSABERGTLALQCEVSDPEAHVVRKDGVOQLGPSDKYDFLHTAGTR 660
 Qy 2702 GLVHDVSPEDAGLYTCHVGSEETRARVRVHDHLVGTITKRLKTNVEVLGESECSPECVLSH 2761
 Db 661 GLVHDVSPEDAGLYTCHVGSEETRARVRVHDHLVGTITKRLKTNVEVLGESECSPECVLSH 720
 Qy 2762 ESADPAMWTVGGKTVGSSSFQATROCKKXILVVRAPSDAGEVVFVRGLTSKASLI 2821
 Db 721 ESADPAMWTVGGKTVGSSSFQATROCKKXILVVRAPSDAGEVVFVRGLTSKASLI 780
 Qy 2822 VRERPAALIKPLEDOWAPGDEVELRCELSRAGTPVHMLKDKAIRKSQKYDVVCEGTMA 2881
 Db 781 VRERPAALIKPLEDOWAPGDEVELRCELSRAGTPVHMLKDKAIRKSQKYDVVCEGTMA 840
 Qy 2882 MLVIRGASLKDAGEYTCVEASKSTASLHVEEKANCETEELTNLOVEKGTAVFTCKTEH 2941
 Db 841 MLVIRGASLKDAGEYTCVEASKSTASLHVEEKANCETEELTNLOVEKGTAVFTCKTEH 900
 Qy 2942 PAATVVRKGLLELRASGHQPSQEGTLRLTISALEKADSDTYTCDIGQAQRAQLLVQ 3001
 Db 901 PAATVVRKGLLELRASGHQPSQEGTLRLTISALEKADSDTYTCDIGQAQRAQLLVQ 960
 Qy 3002 GRRVHIIEDEVDVQEGSSATFCRISPANYEPVHWFDPKTPHANELNEIDAQPGYH 3061
 Db 961 GRRVHIIEDEVDVQEGSSATFCRISPANYEPVHWFDPKTPHANELNEIDAQPGYH 1020
 Qy 3062 VLTLRQALKDGSITTYFAGQORASALRVTEKESVFSRELTDTATITEGEDTLTVCEST 3121
 Db 1021 VLTLRQALKDGSITTYFAGQORASALRVTEKESVFSRELTDTATITEGEDTLTVCEST 1080
 Qy 3122 CDIIMCWTNDKTLRGSAARCOSHEGHRAQLLITGATLQDSGRVYKCEAGACSSIVRVH 3181
 Db 1081 CDIIMCWTNDKTLRGSAARCOSHEGHRAQLLITGATLQDSGRVYKCEAGACSSIVRVH 1140
 Qy 3182 ARPVRFOALKDLEVLGGATLRCVLSSVAAPVKCYGNNVLRPGDKYSLRQEGAMLEL 3241
 Db 1141 ARPVRFOALKDLEVLGGATLRCVLSSVAAPVKCYGNNVLRPGDKYSLRQEGAMLEL 1200
 Qy 3242 VVRNLRPODSGRYSCSGDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCESKTA 3301
 Db 1201 VVRNLRPODSGRYSCSGDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCESKTA 1260
 Qy 3302 PVENRKSETLRDGRVCLRDGAMCELOIRGLAWVDAAEYSCVCGEERTSASLTIRPMP 3361
 Db 1261 PVENRKSETLRDGRVCLRDGAMCELOIRGLAWVDAAEYSCVCGEERTSASLTIRPMP 1320
 Qy 3362 AHFTGRLRHQSIEGATATLRCESKAAAPVWVRKGRSLRDGRHSLRQDGAVCELIQCG 3421
 Db 1321 AHFTGRLRHQSIEGATATLRCESKAAAPVWVRKGRSLRDGRHSLRQDGAVCELIQCG 1380
 Qy 3422 LAVADAGEYSCVCGEERTSATLTVKALPAKETEGLRNEEAVEGATAMLCESLKVAPVEW 3481
 Db 1381 LAVADAGEYSCVCGEERTSATLTVKALPAKETEGLRNEEAVEGATAMLCESLKVAPVEW 1440
 Qy 3482 RKGPENLRDGRVTLRQEGTRCELOICGLAMADAGEYLCVCGOERTSATLTVRALPARFI 3541
 Db 1441 RKGPENLRDGRVTLRQEGTRCELOICGLAMADAGEYLCVCGOERTSATLTVRALPARFI 1500
 Qy 3542 EDVKNQAREGATAVLOCELNSAAAPVWVRKGSSETLRDGRYSLRQDGTKCELOIRGLAMA 3601
 Db 1501 EDVKNQAREGATAVLOCELNSAAAPVWVRKGSSETLRDGRYSLRQDGTKCELOIRGLAMA 1560
 Qy 3602 DTGEYSCVCGOERTSAMLTVRALPIKETEGLRNEEATEGATAVLRCELSKQAPVWVRKKGH 3661
 Db 1561 DTGEYSCVCGOERTSAMLTVRALPIKETEGLRNEEATEGATAVLRCELSKQAPVWVRKKGH 1620
 Qy 3662 ETLRDGRHSLRQDGARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPKSFIEGLR 3721
 Db 1621 ETLRDGRHSLRQDGARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPKSFIEGLR 1680

Qy 3722 NEEATEGDTATLWCELSKAAAPVWVRKKGHETLRDGRHSLRQDGRSCELOIRGLAVVDAGE 3781
 Db 1681 NEEATEGDTATLWCELSKAAAPVWVRKKGHETLRDGRHSLRQDGRSCELOIRGLAVVDAGE 1740
 Qy 3782 YSCVCGOERTSATLTVRALPARFIEDVKNQAREGATAVLOCELNSAAAPVWVRKGSSETLR 3841
 Db 1741 YSCVCGOERTSATLTVRALPARFIEDVKNQAREGATAVLOCELNSAAAPVWVRKGSSETLR 1800
 Qy 3842 GGDYSLSRDGRTRCELOIRGLHGLSVADTGEYSCVCGOERTSATLTVRAPQVPREPLOSQA 3901
 Db 1801 GGDYSLSRDGRTRCELOIRGLHGLSVADTGEYSCVCGOERTSATLTVRALPARTQDLKTEA 1860
 Qy 3902 EEGSTATLCELSSEPTATVWSKGLQLOANGRRPREPLOGCTAEVLQDLQREDTGEYTC 3961
 Db 1861 SEGATATLCELSK-VAPVWVRKGPETLRDGRYSKQDGRCELOIRGLHGLSVADAGEYSC 1919
 Qy 3962 TCGSOATSATLTVTAAPVRFELRELOHQVEDGTAHLCELSRAGASVWVRKGSLOLEPC 4021
 Db 1920 MCGOERTSAMLTVRALPARFTEGLRNEEATEGATATLCELSKAA-APVWVRKGLEALRDG 1978
 Qy 4022 AKYQVQDGAALVLRGVEQEDAGDYTCDTGHTQSMASLSVVRPRPFKTRLOQLEQET 4081
 Db 1979 DKYSLRQDGAVCELOIRGLAMADNGVYSCVCGOERTSATLTVRALPARFIEDMRNQKATE 2038
 Qy 4082 GDIALCCOLSDAESGAVVQWIKESGVELHAGPKYEMRSQGTARELLIHOLEAKDTGEYAC 4141
 Db 2039 GATVTLQCKLRKA---APVWVRKGPNTLKDGRYSKQDGRCELOIRGLVIADAGEYSC 2095
 Qy 4142 VTGGOKTAASLRVTEPVTIVRGLVDAEVTADEVEFCEVSRAGATGVQWCLQGLPLOS 4201
 Db 2096 ICEQERTSATLTVRALPARFIEDVRNHEATEGATAVLOCELNSKAA--VWVRKGSSETLRD 2153
 Qy 4202 NEVTEVAVRDGRIHTLRKGVTPEDAGTVSFHLGNHASSAOLTVRAPVETILEPQOVOL 4261
 Db 2154 GDRYSLR-QDGRTRCELOIRGLAVEDEGEYLCVCGOERTSATLTVRALPARFIDNNTNQA 2212
 Qy 4262 SGGQDASFOCLRSRASGOEARWALGGVPLQANENNDITVEOGTLLHLLTHKVTLEDACTV 4321
 Db 2213 REGATATLCELSKAAAPVE--WRKGRSLRDGRHSLR-QDGAVCELOIRGLAVADAGEY 2269
 Qy 4322 SFHVTCSSQAQLKV 4336
 Db 2270 SCVCGEERTSATLTV 2284

RESULT 7

AAOI5372

ID AAOI5372 standard; protein; 1665 AA.

XX AC AAOI5372;

XX AC AAOI5372;

XX DT 19-SEP-2002 (first entry)

XX DE Human myosin light chain kinase subfamily-related kinase protein.

XX XX

XX KW Human; gene therapy; chromosome 1; kinase protein;

XX KW myosin light chain kinase subfamily; kinase protein-mediated disease;

XX KW transgenic animal.

XX OS Homo sapiens.

XX XX

XX PN WO200240683-A2.

XX PD 23-MAY-2002.

XX XX

XX PF 22-OCT-2001; 2001WO-US032616.

XX XX

XX PR 14-NOV-2000; 2000US-00711134.

XX PR 17-MAY-2001; 2001US-00858664.

XX PA (PEKE) PE CORP NY.

XX XX

Wei M, Ketchum K, Di Francesco V, Beasley EM;

WPI; 2002-500223/53.

N-PSDB; AAL43908, AAL43909.

New kinase proteins related to myosin light chain kinase subfamily and encoding polynucleotide, useful for diagnosing, treating disease or condition mediated by the kinase protein and for identifying modulators.

Claim 1: Fig 2: 96pp: English.

The invention comprises the amino acid and coding sequences (located on chromosome 1) of a human kinase protein that is related to the myosin light chain kinase subfamily. The human kinase DNA and protein sequences of the invention are useful for identifying agents that modulate the activity of the human kinase protein. Kinase-modulating agents are useful for treating a disease or condition mediated by a human kinase protein. The human kinase DNA sequences can be used to produce transgenic animals which are useful for studying the function of kinase proteins and identifying/evaluating modulators of kinase protein activity. The present amino acid sequence represents the human kinase protein of the invention

Sequence 1665 AA:

Query Match 20.4%; Score 8423; DB 5; Length 1665;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 1; Indels 0

| | | | | | |
|------|----|--------------------------|--------------------------|------------------------------|------|
| 6356 | QY | PPSQMVTIEDVQACTGTAQFEAII | EGDPOPSVTWYKDSVQLVDSTRLS | QQOQGGTTYSLV | 6415 |
| 53 | Db | PPSQMVTIEDVQACTGTAQFEAII | EGDPOPSVTWYKDSVQLVDSTRLS | QQOQGGTTYSLV | 112 |
| 6416 | QY | LRHVASKDAGVYTCLAQNTGGQVL | CKAEILLV | LGDNPEPDESEKSHRRKLSHFVVEKEIG | 6475 |
| 113 | Db | LRHVASKDAGVYTCLAQNTGGQVL | CKAEILLV | LGDNPEPDESEKSHRRKLSHFVVEKEIG | 172 |
| 6476 | QY | RGVFGFKVRVQHGNKILCAAKFI | PLRSRTRAQAYRERDILAALS | HPVLVTGLLDQFETRK | 6535 |
| 173 | Db | RGVFGFKVRVQHGNKILCAAKFI | PLRSRTRAQAYRERDILAALS | HPVLVTGLLDQFETRK | 232 |
| 6536 | QY | TLILILELCSSEBLLDRLYKGV | VTAEVAVKY | IQQLVEGLHYLHSHGVTHLHDIKPSNLM | 6595 |
| 233 | Db | TLILILELCSSEBLLDRLYKGV | VTAEVAVKY | IQQLVEGLHYLHSHGVTHLHDIKPSNLM | 292 |
| 6596 | QY | VHPAREDIKICDFGFAQNIITPAE | LQFSQYGSPEFVSP | PEIITQONPVSEASDIWANGVISYL | 6655 |
| 293 | Db | VHPAREDIKICDFGFAQNIITPAE | LQFSQYGSPEFVSP | PEIITQONPVSEASDIWANGVISYL | 352 |
| 6656 | QY | SLTCCSSPFAGESDRATLLNV | LEGRVSWSSPMAAHL | SEDAKDFTKATLQAPQAPPSAAQC | 6715 |
| 353 | Db | SLTCCSSPFAGESDRATLLNV | LEGRVSWSSPMAAHL | SEDAKDFTKATLQAPQAPPSAAQC | 412 |
| 6716 | QY | LSPHPWFLKSPAEAEHPINTKQ | LKFLILARSRWORS | LSYKSIILVMRSIPELLRGPPDPS | 6775 |
| 413 | Db | LSPHPWFLKSPAEAEHPINTKQ | LKFLILARSRWORS | LSYKSIILVMRSIPELLRGPPDPS | 472 |
| 6776 | QY | LGVARHLCRTDGGSSSSSSSD | NELAPFAKASLPPSPV | THSPILLHPRGFIRPSASLPEE | 6835 |
| 473 | Db | LGVARHLCRTDGGSSSSSSSD | NELAPFAKASLPPSPV | THSPILLHPRGFIRPSASLPEE | 532 |
| 6836 | QY | AEASERSTEAPAPPASPEGAG | PPAAQCVPRHSVIRSLF | THQAGESPEHGALAPGSRRHP | 6895 |
| 533 | Db | AEASERSTEAPAPPASPEGAG | PPAAQCVPRHSVIRSLF | THQAGESPEHGALAPGSRRHP | 592 |
| 6896 | QY | ARRRHLKGGYIAGALPGL | REPLMEHRVLBEEAA | REEQATLLAKAPSFETALRIPASGTH | 6955 |
| 593 | Db | ARRRHLKGGYIAGALPGL | REPLMEHRVLBEEAA | REEQATLLAKAPSFETALRIPASGTH | 652 |
| 6956 | QY | LAPCHSHSLHSDSPSTPRP | SEACEAQRULPSAPSG | APIRDMGHGPGSKQLPSTGGHPG | 7015 |
| 653 | Db | LAPCHSHSLHSDSPSTPRP | SEACEAQRULPSAPSG | APIRDMGHGPGSKQLPSTGGHPG | 712 |
| 7016 | QY | TAOPERESPSPWQCAPPFF | CHPKQGSAPQSGCSP | HPHAPVACPFGSGFPFGCKCAPLVPS | 7075 |

Db 953 EVGRAPTRSSPEPTWEDIGQVLSVQIRDSGDAADTISLDISEVDPAYNLSLDYDI 1012
Qy 7316 KYLPFERMI FRKVPKSAQPPSPMAEELAEFPETWMPGELGPHAGLEITEESDND 7375
Db 1013 KYLPFERMI FRKVPKSAQPPSPMAEELAEFPETWMPGELGPHAGLEITEESDND 1072
Qy 7376 ALLAEAAVGRKRWSSPSRSIFHPGSHLPDDEPAELGLRERKVASVEHLSRIILKGRPEG 7435
Db 1073 ALLAEAAVGRKRWSSPSRSIFHPGSHLPDDEPAELGLRERKVASVEHLSRIILKGRPEG 1132
Qy 7436 LEKEGPRKKEGLASFRISGLKSWDRAPTFULRELSDETIVVLGQSVTLACQVSAQPAQAT 7495
Db 1133 LEKEGPRKKEGLASFRISGLKSWDRAPTFULRELSDETIVVLGQSVTLACQVSAQPAQAT 1192
Qy 7496 WSKDGAPELESSRVLISATLKNFOLLILVVAEDLGVYTCVSNALGTVTITGVLRKAE 7555
Db 1193 WSKDGAPELESSRVLISATLKNFOLLILVVAEDLGVYTCVSNALGTVTITGVLRKAE 1252
Qy 7556 RPSSSPCPDICEVADGVLLVWKPVESYGPVTYIVQCSLEGSWTTILASDIFDCCYLTSK 7615
Db 1253 RPSSSPCPDICEVADGVLLVWKPVESYGPVTYIVQCSLEGSWTTILASDIFDCCYLTSK 1312
Qy 7616 LSRGGTYTFRACVSKAGMGPYSFSPBOVLIGGSPSHLASBEESQGRSAQPLPSTKTPAQ 7675
Db 1313 LSRGGTYTFRACVSKAGMGPYSFSPBOVLIGGSPSHLASBEESQGRSAQPLPSTKTPAQ 1372
Qy 7676 TOIORGRFSVVRQCEKASGRALAAKIPYHPKDKTAVLREYALKGLRHPHQAOLHAAY 7735
Db 1373 TOIORGRFSVVRQCEKASGRALAAKIPYHPKDKTAVLREYALKGLRHPHQAOLHAAY 1432
Qy 7736 LSPRHLVLILELSCGPELLCLAPRASYSSEVKDYLMQMLSATQYLHNOHILHLDRSE 7795
Db 1433 LSPRHLVLILELSCGPELLCLAPRASYSSEVKDYLMQMLSATQYLHNOHILHLDRSE 1492
Qy 7796 NMITEYNLLKVDLGNQASLSQKVLPSDFKDYLETMAPELLEGOGAVPQTDIWAIGV 7855
Db 1493 NMITEYNLLKVDLGNQASLSQKVLPSDFKDYLETMAPELLEGOGAVPQTDIWAIGV 1552
Qy 7856 TAFIMLSAEYFVSSEGARDLQGLRKGILVRLSRVCYAGLSGGAFLRSTLCAQPWGRPCA 7915
Db 1553 TAFIMLSAEYFVSSEGARDLQGLRKGILVRLSRVCYAGLSGGAFLRSTLCAQPWGRPCA 1612
Qy 7916 SSCIQCPWLTEEGPACSPAPVTFPTARLRFVVRNREKRALLYKRNHLAQVR 7968
Db 1613 SSCIQCPWLTEEGPACSPAPVTFPTARLRFVVRNREKRALLYKRNHLAQVR 1665

RESULT 9
AAB85504
ID AAB85504 standard; protein; 1618 AA.
XX AC AAB85504;
XX DT 25-SEP-2001 (first entry)
XX DE Human protein kinase SGK145.

Protein kinase; enzyme; cytostatic; neurotropic; neuroprotective; human;
anti-parkinsonian; virucide; antibacterial; antifungal; antimigraine;
analgesic; hypotensive; hypertensive; immunosuppressive; antiallergic;
antipsoriatic; antirheumatic; antiarthritic; ophthalmologic; anorectic;
osteopathic; thrombolytic; antiarteriosclerotic; antiasthmatic;
vasotropic; antidiabetic; gene therapy.
OS Homo sapiens.
XX WO200155356-A2.
XX PN 02-AUG-2001.
XX PD 25-JAN-2001; 2001WO-US002337.
XX PF 301 VISYLSITCSSPAGESDRATLNVLEGRVSSPMAAHLSEDAKDFIKATLQRAQARP 360

PR 25-JAN-2000; 2000US-0178079P.
PR 31-JAN-2000; 2000US-0179364P.
PR 17-FEB-2000; 2000US-0183173P.
PR 17-MAR-2000; 2000US-0190162P.
PR 29-MAR-2000; 2000US-0193404P.
PR 13-NOV-2000; 2000US-0247013P.
XX (SUGB-) SUGEN INC.
PA Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;
XX WPI; 2001-476202/51.
DR N-PSDB; AAH46904.
XX Kinase polypeptides useful for treating cancers, Alzheimer's disease,
PT viral infections, diabetes, obesity, organ transplant rejection and
PT rheumatoid arthritis.
XX Claim 7; Page 215; 218pp; English.
XX The invention provides human protein kinases and protein kinase-like
CC enzymes and polynucleotides encoding the polypeptides. The kinase
CC polypeptides and their modulators are useful for treating a disease or
CC disorder such as cancer, immune-related diseases, cardiovascular disease,
CC brain or neuronal-associated disease and metabolic disorders, including
CC cancers of tissues, cancers of hematopoietic origin, diseases of the
CC central nervous system, diseases of the peripheral nervous system,
CC Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic
CC lateral sclerosis, viral infections, infections caused by prions,
CC bacteria and fungi, ocular diseases, migraines, pain, sexual dysfunction,
CC mood disorders, attention disorders, neurological disorders, hypotension,
CC hypertension, psychotic disorders, neurologic disorders, dyskinesias,
CC metabolic disorders, and organ transplant rejection. They are also useful
CC for treating rhinitis, autoimmunity, atherosclerosis, psoriasis,
CC osteoarthritis, asthma, chronic inflammatory pelvic disease, chronic
CC inflammatory bowel disease, rheumatoid arthritis, metabolic disorders
CC such as diabetes, obesity, cardiovascular diseases such as reperfusion
CC injury, coronary thrombosis, clotting disorders and atherosclerosis.
CC Ocular diseases such as glaucoma, retinopathy and macular degeneration,
CC psychiatric and neurological disorders such as anxiety, schizophrenia,
CC dementia, manic depression, etc. The polynucleotides are useful in gene
CC therapy techniques to treat the above mentioned disorders. Sequences
CC AAB85491-85522 represent the human protein kinases of the invention
XX Sequence 1618 AA;
Qy Query Match 20.4%; Score 8407; DB 4; Length 1618;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 0; Indels 6; Gaps 1;
Qy 6357 PSMQVTIEDVQAQTGTAQFEALIEGDPQPSVTWYKDSVQVLDSTLSQQCGTTVSLVL 6416
Db 1 PSMQVTIEDVQAQTGTAQFEALIEGDPQPSVTWYKDSVQVLDSTLSQQCGTTVSLVL 60
Qy 6417 RHVASKDAGVYTCIAQNTGGQVLCKAELLVLG-----GDNEPDSKQSHRKLHSFYEV 6470
Db 61 RHVASKDAGVYTCIAQNTGGQVLCKAELLVLGAAASHLGDNEPDSKQSHRKLHSFYEV 120
Qy 6471 KEBIGRGVGFVVRVQHKNGKILCAAKFIPLSRTRAQAYRERDILAAISHPLVTLGDDQ 6530
Db 121 KEBIGRGVGFVVRVQHKNGKILCAAKFIPLSRTRAQAYRERDILAAISHPLVTLGDDQ 180
Qy 6531 FETRKTLILILELCSSEELLDRLYRGVVTAEVKYVIOQLVEGLHYLHSHGLHLDIKP 6590
Db 181 FETRKTLILILELCSSEELLDRLYRGVVTAEVKYVIOQLVEGLHYLHSHGLHLDIKP 240
Qy 6591 SNILMVHPAREDIKICDFGFAQNITPAELQFSGYGFVSPEIIOQNVSASDILWAMG 6650
Db 241 SNILMVHPAREDIKICDFGFAQNITPAELQFSGYGFVSPEIIOQNVSASDILWAMG 300
Qy 6651 VISYLSITCSSPAGESDRATLNVLEGRVSSPMAAHLSEDAKDFIKATLQRAQARP 6710
Db 301 VISYLSITCSSPAGESDRATLNVLEGRVSSPMAAHLSEDAKDFIKATLQRAQARP 360

| | | | | | | |
|-----------------------|------|---|---------------|-----------|--------------|--|
| Query Match | | 20.4%; | Score 8403; | DB 4; | Length 1610; | |
| Best Local Similarity | | 99.9%; | Pred. No. 0; | | | |
| Matches 1609; | | Conservative 0; | Mismatches 1; | Indels 0; | Gaps 0; | |
| Qy | 6359 | MQVTIEDVQAGTGTAGTAEIIEGDPQPSVTWKDVSQVLDSTRLSQQQEGTYSVLVRH | 6418 | | | |
| Db | 1 | MQVTIEDVQAGTGTAGTAEIIEGDPQPSVTWKDVSQVLDSTRLSQQQEGTYSVLVRH | 60 | | | |
| Qy | 6419 | VASXDAGVYTCLAQNTGGQVLCKAELLVGLGDNEPDESEKSHRKHLSHFVEVEEIGRGV | 6478 | | | |
| Db | 61 | VASXDAGVYTCLAQNTGGQVLCKAELLVGLGDNEPDESEKSHRKHLSHFVEVEEIGRGV | 120 | | | |
| Qy | 6479 | FGFVYRQVHKGNTKILCAAKTIPLRSRTRAQAYRERDILAALSHPLVTGLLDQFETRTKLI | 6538 | | | |
| Db | 121 | FGFVYRQVHKGNTKILCAAKTIPLRSRTRAQAYRERDILAALSHPLVTGLLDQFETRTKLI | 180 | | | |
| Qy | 6539 | LILELCSSEELLRLYRKGVTEAEVXYIQQLVGGLHYLHSHCVLHLDIKPSNIMLVHP | 6598 | | | |
| Db | 181 | LILELCSSEELLRLYRKGVTEAEVXYIQQLVGGLHYLHSHCVLHLDIKPSNIMLVHP | 240 | | | |
| Qy | 6599 | AREDIKICDFGAQNIIPAEILQFQSGSPFVSPETIQONPVSEASDIWANGVISYLSLT | 6658 | | | |
| Db | 241 | AREDIKICDFGAQNIIPAEILQFQSGSPFVSPETIQONPVSEASDIWANGVISYLSLT | 300 | | | |
| Qy | 6659 | CSSPFAGESDRATLNLVLEGRVSWSPMAHLSDAKDFIKATLORAPQAPRAAQCCLSH | 6718 | | | |
| Db | 301 | CSSPFAGESDRATLNLVLEGRVSWSPMAHLSDAKDFIKATLORAPQAPRAAQCCLSH | 360 | | | |
| Qy | 6719 | PWFLKSPAEBAHINTKQLKFLARQWORSLSYKSIILVMSIPELLRGPDSPSLGV | 6778 | | | |
| Db | 361 | PWFLKSPAEBAHINTKQLKFLARQWORSLSYKSIILVMSIPELLRGPDSPSLGV | 420 | | | |
| Qy | 6779 | ARHLCRTDGGSSSSSSSDNELAPFAKASLPSPSVTHSLPHRPGFLRPSASLPBEAEA | 6838 | | | |
| Db | 421 | ARHLCRTDGGSSSSSSSDNELAPFAKASLPSPSVTHSLPHRPGFLRPSASLPBEAEA | 480 | | | |
| Qy | 6839 | SERSTEAPAPASPEGAPPAQCGVPRHSVIRSLFYHQGESPEHGALAPGRRHPARR | 6898 | | | |
| Db | 481 | SERSTEAPAPASPEGAPPAQCGVPRHSVIRSLFYHQGESPEHGALAPGRRHPARR | 540 | | | |
| Qy | 6899 | RHLKGGYIAGALPGLREPLMEHRLVLEEAAAREEQATLLAKAPSFETALRLPASGTHLAP | 6958 | | | |
| Db | 541 | RHLKGGYIAGALPGLREPLMEHRLVLEEAAAREEQATLLAKAPSFETALRLPASGTHLAP | 600 | | | |
| Qy | 6959 | GHSHLSHDSSTPRPSEAGEAORLPSAPSGAPIRDMCHPQSGKOLSTGCHPCTAQ | 7018 | | | |
| Db | 601 | GHSHLSHDSSTPRPSEAGEAORLPSAPSGAPIRDMCHPQSGKOLSTGCHPCTAQ | 660 | | | |
| Qy | 7019 | PERSPSPWQAPAPFCHPKQGSAPQGCSPHPAVAPCPGSPFPFGCKEAPLVPSSPFL | 7078 | | | |
| Db | 661 | PERSPSPWQAPAPFCHPKQGSAPQGCSPHPAVAPCPGSPFPFGCKEAPLVPSSPFL | 720 | | | |
| Qy | 7079 | GQOAPAPAKASPLDQKMGPGDISLPGRPKPCSSPGSASQASSQVSSLRVGSQV | 7138 | | | |
| Db | 721 | GQOAPAPAKASPLDQKMGPGDISLPGRPKPCSSPGSASQASSQVSSLRVGSQV | 780 | | | |
| Qy | 7139 | GTEPGPSLDAGWTQAEEDLSDTPTLQRPQEQATMEKFSILGGSGVAGVAGTTFAPGG | 7198 | | | |
| Db | 781 | GTEPGPSLDAGWTQAEEDLSDTPTLQRPQEQATMEKFSILGGSGVAGVAGTTFAPGG | 840 | | | |
| Qy | 7199 | DAGMGLGQGPWARIAMAVSQSEEEQEPARAEQSEEQEPARAEPLQVQARPVEVG | 7258 | | | |
| Db | 841 | DAGMGLGQGPWARIAMAVSQSEEEQEPARAEQSEEQEPARAEPLQVQARPVEVG | 900 | | | |
| Qy | 7259 | RAPTRSSPEPTPWEIDIGVSLVQIRDLSDGAERADTISLDISEVDPAVLNLSLDYDKYL | 7318 | | | |
| Db | 901 | RAPTRSSPEPTPWEIDIGVSLVQIRDLSDGAERADTISLDISEVDPAVLNLSLDYDKYL | 960 | | | |
| Qy | 7319 | PFEFMIFRKVPKSAQPEPPSPWABEELAEFPETWMPGELGPHAGLEITEESDVEDDALL | 7378 | | | |
| Db | 961 | PFEFMIFRKVPKSAQPEPPSPWABEELAEFPETWMPGELGPHAGLEITEESDVEDDALL | 1020 | | | |
| Qy | 7379 | AEAAVGRKRWKSSPSSRLTFHPGRLHPLDEPAELGLRERVKASVEHISRLILKGRPEGLEK | 7438 | | | |

| | | | | | | |
|----|------|---|------|--|--|--|
| Db | 1021 | AEAAVGRKRWKSSPSSRLTFHPGRLHPLDEPAELGLRERVKASVEHISRLILKGRPEGLEK | 1080 | | | |
| Qy | 7439 | EGPPKKKGLASFRLSGLKSWDRAPTFELRELSDETVVLGQSVTLACQVSAQAPAAQATWSK | 7498 | | | |
| Db | 1081 | EGPPKKKGLASFRLSGLKSWDRAPTFELRELSDETVVLGQSVTLACQVSAQAPAAQATWSK | 1140 | | | |
| Qy | 7499 | DCAPLESSSRVLISATLKNFQLLTILVVVAEDLGYITCSVSNALGTVTTTGVLRKAERPS | 7558 | | | |
| Db | 1141 | DCAPLESSSRVLISATLKNFQLLTILVVVAEDLGYITCSVSNALGTVTTTGVLRKAERPS | 1200 | | | |
| Qy | 7559 | SSPCPDIGEVYADGVLLWKPVESYGPVTYIVQCSLEGSSWTTLASDIFDCCLTSLKLSR | 7618 | | | |
| Db | 1201 | SSPCPDIGEVYADGVLLWKPVESYGPVTYIVQCSLEGSSWTTLASDIFDCCLTSLKLSR | 1260 | | | |
| Qy | 7619 | GTGTYTFRCTACVSKAGMGPSSESPSEQVLLGAPSHLASEESQGRSAQPLPSIKTFAFQOI | 7678 | | | |
| Db | 1261 | GTGTYTFRCTACVSKAGMGPSSESPSEQVLLGAPSHLASEESQGRSAQPLPSIKTFAFQOI | 1320 | | | |
| Qy | 7679 | QGRFSVVRQCVKEKASGRALAAKII:PYHPKOKTAVLREYEAALKGLRPHLAQLHAAVYLS | 7738 | | | |
| Db | 1321 | QGRFSVVRQCVKEKASGRALAAKII:PYHPKOKTAVLREYEAALKGLRPHLAQLHAAVYLS | 1380 | | | |
| Qy | 7739 | RHLVILELCSGPELLPCLABEASYSSEVKDYLMQMLSATOYLHNOHTLHLDLSENMI | 7798 | | | |
| Db | 1381 | RHLVILELCSGPELLPCLABEASYSSEVKDYLMQMLSATOYLHNOHTLHLDLSENMI | 1440 | | | |
| Qy | 7799 | ITEYNLLKVDLGNAGSLSQEKVLPDKFYLETMAPELLEGGQAVPQTDIWAIGVTA | 7858 | | | |
| Db | 1441 | ITEYNLLKVDLGNAGSLSQEKVLPDKFYLETMAPELLEGGQAVPQTDIWAIGVTA | 1500 | | | |
| Qy | 7859 | IMLSAEYPVSSSGARDLQRLKGLVRLSRCVAGLSGGAVAFRLSTLCAQPKGRPCASS | 7918 | | | |
| Db | 1501 | IMLSAEYPVSSSGARDLQRLKGLVRLSRCVAGLSGGAVAFRLSTLCAQPKGRPCASS | 1560 | | | |
| Qy | 7919 | LQCPWLTEBGPACSRPAPVTFTFARLVFVRNRRERALLYKRNHLAQVR | 7968 | | | |
| Db | 1561 | LQCPWLTEBGPACSRPAPVTFTFARLVFVRNRRERALLYKRNHLAQVR | 1610 | | | |

RESULT 11

AAB30567

ID AAB30567 standard; protein; 1351 AA.

XX AC AAB30567;

XX DT 19-MAR-2001 (first entry)

XX DE Amino acid sequence of a human signal transduction polypeptide.

XX KW Signal transduction; H19C5; kinase; cardiac disease; angina pectoris;

XX KW congestive heart failure; dilated congestive cardiomyopathy;

XX KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;

XX KW mitral valve disease; aortic valve disease; tricuspid valve disease;

XX KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;

XX KW atherosclerosis; cardiac tumour; microbial infection.

XX OS Homo sapiens.

XX PN WO200063381-A1.

XX PD 26-OCT-2000.

XX PF 11-APR-2000; 2000WO-US0009488.

XX PR 16-APR-1999; 99US-0129553P.

XX PA (SCIO-) SCIOS INC.

XX Zeng W, Stanton L, Kong H;

XX WPI; 2001-007013/01.

XX N-PSDB; AAC62285.

XX Novel h19G5 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
XX cardiac diseases, and additional mediators of signal transduction.
PS Claim 1; Page 55-57; 81pp; English.
XX The present sequence represents a human protein with putative function in
CC signal transduction. The polypeptide is designated H19G5. The protein is
CC capable of regulating signal transduction and exhibits kinase activity.
CC The H19G5 transcript is expressed in the heart. H19G5 polypeptides and
CC polynucleotides are useful for preventing or treating a cardiac disease,
CC such as congestive heart failure, dilated congestive cardiomyopathy,
CC hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve
CC disease, aortic valve disease or tricuspid valve disease, angina
CC pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial
CC or renovascular hypertension, arteriosclerosis, atherosclerosis and
CC cardiac tumours in humans. The polypeptide is also useful for detecting
CC the expression of a protein capable of regulating signal transduction or
CC the expression of a protein capable of acting as a donor or acceptor
CC molecule of a phosphate group. The monoclonal antibodies can be used as
CC probes for detecting discrete antigens expressed by tissue or cell
CC samples, and therefore used in humans for localization and monitoring of
CC microbial infection
XX Sequence 1351 AA;

Query Match 17.1%; Score 7073; DB 4; Length 1351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1349; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6619 LQFQYQSPFVPSVEIIQQNPVPSASDIWANGVLSYLSLTCSPFFAGESDRAILLNVLG 6678
DB 2 VQFQYQSPFVPSVEIIQQNPVPSASDIWANGVLSYLSLTCSPFFAGESDRAILLNVLG 61

QY 6679 RVSSSPMAHLSDAKDFIKATLQAPQAPRPAQAACLSHPWFLKSNPAEAEHINTKQL 6738
DB 62 RVSSSPMAHLSDAKDFIKATLQAPQAPRPAQAACLSHPWFLKSNPAEAEHINTKQL 121

QY 6739 KFLARSRWORSLSYKSLVMSIPPELLRGPDPSLGVARHLCRDTGGSSSSSSSSN 6798
DB 122 KFLARSRWORSLSYKSLVMSIPPELLRGPDPSLGVARHLCRDTGGSSSSSSSSN 181

QY 6799 ELAPFARAKSLPPSPVTHSLPHPRGFLRPSASLPPEAEASERSTEAPAPPSPGAGPP 6858
DB 182 ELAPFARAKSLPPSPVTHSLPHPRGFLRPSASLPPEAEASERSTEAPAPPSPGAGPP 241

QY 6859 AAQCQVPRHSVIRSLFVHQGESPEHGAALAPGSRHPARRRHLLKGYTAGALPGIREFL 6918
DB 242 AAQCQVPRHSVIRSLFVHQGESPEHGAALAPGSRHPARRRHLLKGYTAGALPGIREFL 301

QY 6919 MEHVLSEEAAREQAALLAKAPSFETALRLPASGTHLAPGSHSLHSDSPSPRPSSEA 6978
DB 302 MEHVLSEEAAREQAALLAKAPSFETALRLPASGTHLAPGSHSLHSDSPSPRPSSEA 361

QY 6979 CGEQRILPSAPSGAPTRDMCHPGQSLPSTGCHPGTAQPERPSPDPSWGQAPFCHPK 7038
DB 362 CGEQRILPSAPSGAPTRDMCHPGQSLPSTGCHPGTAQPERPSPDPSWGQAPFCHPK 421

QY 7039 QGSAPOGCGPHPAVAPCPGSPFPQCKEAPLVPSPPFLGQAPQAPPAKASFPPLDSKM 7098
DB 422 QGSAPOGCGPHPAVAPCPGSPFPQCKEAPLVPSPPFLGQAPQAPPAKASFPPLDSKM 481

QY 7099 GPGDISILPGKPKGPCSSPGSASQASSQVSSIRLVGSSQVGTGTPGSPSLDAEGWTQAEADL 7158
DB 482 GPGDISILPGKPKGPCSSPGSASQASSQVSSIRLVGSSQVGTGTPGSPSLDAEGWTQAEADL 541

QY 7159 SDSPTLQRPQEQATMKFKSLGGRGGYAGVAGTGFAGGDAGMLGQGMWARIAWAYS 7218
DB 542 SDSPTLQRPQEQATMKFKSLGGRGGYAGVAGTGFAGGDAGMLGQGMWARIAWAYS 601

QY 7219 QSEEEQEAEARASQSEEQAEARASPLPQVSARVPVEVGRAPTRSPPEPTPWEDIGQVS 7278

DB 602 QSEEEQEAEARASQSEEQAEARASPLPQVSARVPVEVGRAPTRSPPEPTPWEDIGQVS 661
QY 7279 LVQIRDLSDGDAEAADTISLDISEVDPAVNLSDLYDIKYLPEFFEMIFRKPVKSAQPPPS 7338
DB 662 LVQIRDLSDGDAEAADTISLDISEVDPAVNLSDLYDIKYLPEFFEMIFRKPVKSAQPPPS 721
QY 7339 PMAEELAEFFPEPTWMPGELGPHAGLBITESSEVDALLAEAAVGRKRKWSRSRSLFH 7398
DB 722 PMAEELAEFFPEPTWMPGELGPHAGLBITESSEVDALLAEAAVGRKRKWSRSRSLFH 781
QY 7399 PFRHLPLDEPAELGLRERVVASVEHISRIUKGPEGLKEGPKRKPGLASFRSLGSKS 7458
DB 782 PFRHLPLDEPAELGLRERVVASVEHISRIUKGPEGLKEGPKRKPGLASFRSLGSKS 841
QY 7459 WDRAPTFLRELSDETIVTIGQSVTLACQVSAQAAQATMSKDGAPLESRRVLIISATLKNF 7518
DB 842 WDRAPTFLRELSDETIVTIGQSVTLACQVSAQAAQATMSKDGAPLESRRVLIISATLKNF 901
QY 7519 QLLTILVVAEDLGVYTCVSNALGTVTTTGVLSKAEPPSSPCPDIGEVADGVLVWK 7578
DB 902 QLLTILVVAEDLGVYTCVSNALGTVTTTGVLSKAEPPSSPCPDIGEVADGVLVWK 961
QY 7579 PVESYGPVTYIVQCSLEGSSWTTLASDIFDCCYLTSLSRGTYTFTTACVSKAGMGFYS 7638
DB 962 PVESYGPVTYIVQCSLEGSSWTTLASDIFDCCYLTSLSRGTYTFTTACVSKAGMGFYS 1021

QY 7639 SPSEQVILGGPSHLASEEESQGRSAQPLPSTKTAFQTOIGRFRSVVRCWEKASGRAL 7698
DB 1022 SPSEQVILGGPSHLASEEESQGRSAQPLPSTKTAFQTOIGRFRSVVRCWEKASGRAL 1081

QY 7699 AAKIIPYHPKDKTAVLREYEALKGLRHPLAQLHAAVLSRPHVLIIELCSPPELLPCLA 7758
DB 1082 AAKIIPYHPKDKTAVLREYEALKGLRHPLAQLHAAVLSRPHVLIIELCSPPELLPCLA 1141

QY 7759 BRASYSESEVKDYLMQMLSATQYLHNOHILHLDLRSENMIITEYNLLKVDLGNASLSQ 7818
DB 1142 BRASYSESEVKDYLMQMLSATQYLHNOHILHLDLRSENMIITEYNLLKVDLGNASLSQ 1201

QY 7819 EKVLPSPDFKDYLETMAPELLEGQAVPQTDIWAIGVTAFTMLSAEYVPSSEGARDLQRG 7878
DB 1202 EKVLPSPDFKDYLETMAPELLEGQAVPQTDIWAIGVTAFTMLSAEYVPSSEGARDLQRG 1261

QY 7879 LRKGLVRLSRVYAGLSGGAFLSTLCAQWGRPCASSCLQCPWLTEBGPACSRPAPVT 7938
DB 1262 LRKGLVRLSRVYAGLSGGAFLSTLCAQWGRPCASSCLQCPWLTEBGPACSRPAPVT 1321

QY 7939 PPTARLVFVNRERKRALYKRNLAQVR 7968
DB 1322 PPTARLVFVNRERKRALYKRNLAQVR 1351

RESULT 12

AAB30570

ID AAB30570 standard; protein; 871 AA.

XX AAB30570;

XX AC AAB30570;

XX DT 19-MAR-2001 (first entry)

XX A splice variant of a signal transduction polypeptide.

DE Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
KW congestive heart failure; dilated congestive cardiomyopathy;
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
KW mitral valve disease; aortic valve disease; tricuspid valve disease;
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;
KW atherosclerosis; cardiac tumour; microbial infection; splice variant.

XX Homo sapiens.

XX OS WO2000063381-A1.

XX XX 26-OCT-2000.

XX PD

XX 11-APR-2000; 2000WO-US009488.
PF 16-APR-1999; 99US-0129553P.
XX (SCIO-) SCIOS INC.
XX Zeng W, Stanton L, Kong H;
PI WPI; 2001-007013/01.
XX Novel h19G5 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
PT cardiac diseases, and additional mediators of signal transduction.
XX
PS Claim 1; Page 74-76; 81pp; English.
CC The present sequence represents a splice variant of human in signal
CC transduction polypeptide. The polypeptide is designated H19G5. The
CC protein is capable of regulating signal transduction and exhibits kinase
CC activity. The H19G5 transcript is expressed in the heart. H19G5
CC polypeptides and polynucleotides are useful for preventing or treating a
CC cardiac disease, such as congestive heart failure, dilated congestive
CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
CC mitral valve disease, aortic valve disease or tricuspid valve disease,
CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
CC arterial or thenovascular hypertension, arteriosclerosis, atherosclerosis
CC and cardiac tumours in humans. The polypeptide is also useful for
CC detecting the expression of a protein capable of regulating signal
CC transduction or the expression of a protein capable of acting as a donor
CC or acceptor molecule of a phosphate group. The monoclonal antibodies can
CC be used as probes for detecting discrete antigens expressed by tissue or
CC cell samples, and therefore used in humans for localization and
CC monitoring of microbial infection
XX
SQ Sequence 871 AA;

Query Match 11.0%; Score 4533; DB 4; Length 871;
Best Local Similarity 100.0%; Pred. No. 2.8e-226; Mismatches 0; Gaps 0;
Matches 871; Conservative 0;

QY 7098 MGPGLISLPGRKPGPCSSPGSASQASSQSSVSLRVGSSQVGTTPGSLDAEGTQBAED 7157
DB 1 MGPGLISLPGRKPGPCSSPGSASQASSQSSVSLRVGSSQVGTTPGSLDAEGTQBAED 60
QY 7158 LSDSTPTLQRFQEQATMRKFSLGGRGGYAGVAGYGTTFAGDAGGMLQGGPMWARIWAV 7217
DB 61 LSDSTPTLQRFQEQATMRKFSLGGRGGYAGVAGYGTTFAGDAGGMLQGGPMWARIWAV 120
QY 7218 SQSEEEQEEARAEQSQSEEQEAREASPLPOVSARPVPEVGRAPTRSSPEPTPWEDIGQV 7277
DB 121 SQSEEEQEEARAEQSQSEEQEAREASPLPOVSARPVPEVGRAPTRSSPEPTPWEDIGQV 180
QY 7278 SLVQIRDLSGDAEAAADTISLDISEVDPAYLNLSDLYDIKYLPPFFPMIPRKVPKSAQBPBP 7337
DB 181 SLVQIRDLSGDAEAAADTISLDISEVDPAYLNLSDLYDIKYLPPFFPMIPRKVPKSAQBPBP 240
QY 7338 SPMAEEELAEPEPTWMPGELGPHAGLEITHEESDVEDALLAEAVGRKRWSSPSRSLF 7397
DB 241 SPMAEEELAEPEPTWMPGELGPHAGLEITHEESDVEDALLAEAVGRKRWSSPSRSLF 300
QY 7398 HFPGRHLPLDEPAELGLRERVKASVEHISRLKGRPEGLEKEGPPRPKPGLASFRLSLGK 7457
DB 301 HFPGRHLPLDEPAELGLRERVKASVEHISRLKGRPEGLEKEGPPRPKPGLASFRLSLGK 360
QY 7458 SMDRAPTFLEISDETAVLGOSVTLACQVSAQPAQAQWTKDGAFLSSSVLISATLKN 7517
DB 361 SMDRAPTFLEISDETAVLGOSVTLACQVSAQPAQAQWTKDGAFLSSSVLISATLKN 420
QY 7518 FOLLITLVVAEDLGVYTCVSNALGTVTTTGLVLRKAERPSSPCPDICEVYADGVLLVW 7577
DB 421 FOLLITLVVAEDLGVYTCVSNALGTVTTTGLVLRKAERPSSPCPDICEVYADGVLLVW 480

QY 7578 KPVESYGPVTYIVQCSLEGGSWTTILASDIFDCCYLTSLKSRGGTYTFTTACVSKAGMGYPY 7637
DB 481 KPVESYGPVTYIVQCSLEGGSWTTILASDIFDCCYLTSLKSRGGTYTFTTACVSKAGMGYPY 540
QY 7638 SSPSEQVILGGPSHLASEEESQGRSAQPLPSTKTAFQTOIGRGSFVVQRQWEXASGRA 7697
DB 541 SSPSEQVILGGPSHLASEEESQGRSAQPLPSTKTAFQTOIGRGSFVVQRQWEXASGRA 600
QY 7698 LAAKIIPYHPKDKTAVLREYEALKGLRHPHLAQLHAAYLSRPHLVLIILELCSGPPELLPCL 7757
DB 601 LAAKIIPYHPKDKTAVLREYEALKGLRHPHLAQLHAAYLSRPHLVLIILELCSGPPELLPCL 660
QY 7758 AERASVSESEVDKYLWQMLSATQYLNQHILHLDRSENMIITEYNLLKVVDLGNAQSLS 7817
DB 661 AERASVSESEVDKYLWQMLSATQYLNQHILHLDRSENMIITEYNLLKVVDLGNAQSLS 720
QY 7818 QEKVLPSDFKDYLETMAPELLEGQAGVQPTDIWAIGVTAFTMLSAEYFVSSEGADLQOR 7877
DB 721 QEKVLPSDFKDYLETMAPELLEGQAGVQPTDIWAIGVTAFTMLSAEYFVSSEGADLQOR 780
QY 7878 GLRKGVLVRLSRVAGLSGGAVAFRLSTLCAQWGRPCASSCLOCPLWTEEGPACSRPAPV 7937
DB 781 GLRKGVLVRLSRVAGLSGGAVAFRLSTLCAQWGRPCASSCLOCPLWTEEGPACSRPAPV 840
QY 7938 TPTPTARLVFVRNREKRALLYKRHNLAQVR 7968
DB 841 TPTPTARLVFVRNREKRALLYKRHNLAQVR 871
RESULT 13
AAE16274
ID AAE16274 standard; protein; 871 AA.
XX AAE16274;
XX 26-MAR-2002 (first entry)
XX Human kinase PKIN-20 protein.
XX Human; kinase; PKIN-20; cancer; leukaemia; adenocarcinoma; osteoporosis;
KW immune disorder; atherosclerosis; Crohn's disease; Hodgkin's disease;
KW Acquired Immune Deficiency Syndrome; AIDS; Addison's disease; anaemia;
KW allergy; asthma; adult respiratory distress syndrome; multiple sclerosis;
KW autoimmune thyroiditis; bronchitis; diabetes mellitus; osteoarthritis;
KW Good pasture's syndrome; Graves' disease; pancreatitis; psoriasis;
KW rheumatoid arthritis; ulcerative colitis; cirrhosis; Cushing's syndrome;
KW hepatitis; hypothyroidism; cerebral palsy; cataract; angina pectoris;
KW cardiovascular disease; hypertension; vasculitis; myocarditis; obesity;
KW congestive heart failure; ischaemic heart disease; lung tumour; gout;
KW fatty liver; Niemann-Pick's disease; gene therapy.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT Domain 575..827
FT FT /note= "Eukaryotic protein kinase domain"
FT Domain 580..812
FT FT /label= Protein_kinase_domain
XX WO200196547-A2.
XX
XX 20-DEC-2001.
XX
XX 14-JUN-2001; 2001WO-US019444.
XX
XX 15-JUN-2000; 2000US-0212073P.
XX 23-JUN-2000; 2000US-0213467P.
XX 30-JUN-2000; 2000US-0215651P.
XX 07-JUL-2000; 2000US-0216605P.
XX 13-JUL-2000; 2000US-0218372P.
XX 25-AUG-2000; 2000US-0228056P.
XX (INCY-) INCYTE GENOMICS INC.
PA

XX Yue H, Lal P, Bandman O, Borowsky ML, Au-Young J, Lu Y;
 PI Gandhi AR, Tribouley CM, Wallia NK, Yao MG, Lu DM, Greenwald SR;
 PI Ramkumar J, Griffin JA, Kearney L, Burford N, Nguyen DB, Tang YT;
 PI Baughn MR, He A, Thornton M, Hafalia A, Patterson C, Gururajan R;
 PI Lo TP, Khan F, Recipon SA, Azimzai Y, Pollick JL, Ding L;
 PI Grether M, Elliott VS, Thangavelu K, Batra S, Ison CH;
 XX WPI; 2002-090207/12.
 DR N-PSDB; AAD26467.
 XX
 PT New polypeptides, useful for diagnosing, treating or preventing disorders
 PT of growth and development, cardiovascular and lipid, and diseases such as
 PT cancer, comprise human kinase polypeptides.
 XX
 PS Claim 1; Page 164-165; 19pp; English.
 XX
 CC The invention relates to human kinase PKIN proteins and their
 CC corresponding cDNAs. A composition containing PKIN agonist is useful for
 CC treating a disease or condition associated with decreased expression of
 CC PKIN and a composition comprising PKIN antagonist is useful for treating
 CC a disease or condition associated with overexpression of PKIN. The
 CC disorders include cancer (leukaemia, adenocarcinoma, lymphoma, melanoma,
 CC myeloma, sarcoma, teratocarcinoma, Hodgkin's disease); immune disorder
 CC (Acquired Immune Deficiency Syndrome (AIDS), asthma, Addison's disease,
 CC atherosclerosis, anaemia, allergies, adult respiratory distress syndrome,
 CC autoimmune thyroiditis, gout, bronchitis, Crohn's disease, diabetes
 CC mellitus, multiple sclerosis, Good pasture's syndrome, Graves' disease,
 CC osteoarthritis, osteoporosis, pancreatitis, psoriasis, Reiter's syndrome,
 CC rheumatoid arthritis, Sjogren's syndrome, uveitis, ulcerative colitis,
 CC bacterial, parasitic, fungal, viral, protozoal and helminthic infections)
 CC growth and development disorders (arteriosclerosis, cirrhosis, hepatitis,
 CC Cushing's syndrome, hypothyroidism, cerebral palsy, cataracts); cardiac
 CC vascular disease (arteriovenous fistula, hypertension, vasculitis,
 CC aneurysms, congestive heart failure, angina pectoris, myocarditis,
 CC ischaemic heart disease, chronic bronchitis, lung tumours); lipid
 CC disorder (fatty liver, Fabry's disease, Niemann-Pick's disease,
 CC hypcholesterolaemia, obesity). PKIN DNA is useful for assessing toxicity
 CC of a test compound and in gene therapy. The present sequence is human
 CC PKIN-20 protein
 XX
 SQ Sequence 871 AA;
 Query Match 11.0%; Score 4529; DB 5; Length 871;
 Best Local Similarity 99.9%; Pred. No. 4.5e-226;
 Matches 870; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 7098 MGPDGSLPGPKPGPCSSPGSASQASSQVSLRVGSSQVGTGPGSLDAEGWTQASD 7157
 DB 1 MGPDGSLPGPKPGPCSSPGSASQASSQVSLRVGSSQVGTGPGSLDAEGWTQASD 60
 QY 7158 LSDSTPTLQRPQEQATWRKSLGGRGYAGVAGTGFAGGAGGMLGQGPWARIANAV 7217
 DB 61 LSDSTPTLQRPQEQATWRKSLGGRGYAGVAGTGFAGGAGGMLGQGPWARIANAV 120
 QY 7218 SQSEEEQEARASQSEEQEARASPLQVARSVPVEVGRAPTRSSPEPTPWEDIGQV 7277
 DB 121 SQSEEEQEARASQSEEQEARASPLQVARSVPVEVGRAPTRSSPEPTPWEDIGQV 180
 QY 7278 SLVQIRLSDGDAADTISLDSVDPAVNLSDLYDKYLPFEFMRKVPKSAQPEPP 7337
 DB 181 SLVQIRLSDGDAADTISLDSVDPAVNLSDLYDKYLPFEFMRKVPKSAQPEPP 240
 QY 7338 SPNAEEELAFPEPTFWPGLGPHAGLEITEESVDALLAAVGRKRWSSPSRSLF 7397
 DB 241 SPNAEEELAFPEPTFWPGLGPHAGLEITEESVDALLAAVGRKRWSSPSRSLF 300
 QY 7398 HFCRHLPLDEPAELGRERVKASVEHISRLKGRPEGLEKEGPPKKKCLASFRLSGLK 7457
 DB 301 HFCRHLPLDEPAELGRERVKASVEHISRLKGRPEGLEKEGPPKKKCLASFRLSGLK 360
 QY 7458 SWDRAPTFLRELSDETVVLGQSVTLACQVSAQAATWSKDGAPLESSRVLISATLKN 7517

DB 361 SWDRAPTFLRELSDETVVLGQSVTLACQVSAQAATWSKDGAPLESSRVLISATLKN 420
 QY 7518 FOLLITLVVVAEDLGVYTCVSYNALGTVTTTGVLRKAERPSPPCPDIGEVYADGVLLAV 7577
 DB 421 FOLLITLVVVAEDLGVYTCVSYNALGTVTTTGVLRKAERPSPPCPDIGEVYADGVLLAV 480
 QY 7578 KEVESYGPVTYIVQCSLEGGSWTTLASDIFDCCYLTSKLSRGGTVTFRTACVSKAGMGFP 7637
 DB 481 KEVESYGPVTYIVQCSLEGGSWTTLASDIFDCCYLTSKLSRGGTVTFRTACVSKAGMGFP 540
 QY 7638 SPSQVLLGGPSHLASBEESQGRSAQPLPSTKTFAPQTIQGRFVSVROCKEASGRA 7697
 DB 541 SPSQVLLGGPSHLASBEESQGRSAQPLPSTKTFAPQTIQGRFVSVROCKEASGRA 600
 QY 7698 LAAKIIPYHPKDKTAVLREYEALKGLRHPHLAQHAAVLSPRHLVLIILELCSGPELLPCL 7757
 DB 601 LAAKIIPYHPKDKTAVLREYEALKGLRHPHLAQHAAVLSPRHLVLIILELCSGPELLPCL 660
 QY 7758 ABRASYSESEVKDYLMQWLSATQYLHNQHLHLDRSENMIITEYNLLKVVLDLGNASLS 7817
 DB 661 ABRASYSESEVKDYLMQWLSATQYLHNQHLHLDRSENMIITEYNLLKVVLDLGNASLS 720
 QY 7818 QEKVLPSDKFKDYLETMAPELLEGQAVPQTDIWAIGVTAIIMLSAEYFVSSEGARDLQR 7877
 DB 721 QEKVLPSDKFKDYLETMAPELLEGQAVPQTDIWAIGVTAIIMLSAEYFVSSEGARDLQR 780
 QY 7878 GLRKLGLVLSRCYAGLSGGAFAFLRSTLCAQPWGRPCASSCLOCPWLTTEGPACSRPAPV 7937
 DB 781 GLRKLGLVLSRCYAGLSGGAFAFLRSTLCAQPWGRPCASSCLOCPWLTTEGPACSRPAPV 840
 QY 7938 TPTTARLVFVRNREKRALLYKRNLAQVR 7968
 DB 841 TPTTARLVFVRNREKRALLYKRNLAQVR 871
 RESULT 14
 ABG74786
 ID ABG74786 standard; protein; 31267 AA.
 XX AC ABG74786;
 XX DT 05-JUN-2003 (first entry)
 XX DE Human RGS11 protein.
 XX KW RGS11; human; screening; cardiant; antianginal; gene therapy;
 XX OS heart disorder; cardiac ischaemia; heart failure; angina.
 XX OS Homo sapiens.
 XX PN WO2002103355-A1.
 XX PD 27-DEC-2002.
 XX PF 17-JUN-2002; 2002WO-JP006019.
 XX PR 18-JUN-2001; 2001JP-00183038.
 XX PA (TAKE) TAKEDA CHEM IND LTD.
 XX PI Koyama N, Tanida S, Yamamoto K;
 XX DR WPI; 2003-167557/16.
 XX PT N-PSDB; ABX113540.
 XX PT Screening compounds regulating RGS11 expression and activity for
 XX prevention and treatment of heart disease.
 XX PS Claim 1; Page 59-261; 321pp; Japanese.
 CC This invention describes a novel method for screening compounds for their
 CC ability to regulate the activity and expression of human RGS11 and its
 CC partial peptides and salts, by observing the expression or activity of

CC RGS11 in the presence or absence of the test compound. The products of
CC the invention have cardiant and antiangiogenic activity and can be used for
CC gene therapy. The methods and compositions are useful in the prevention,
CC treatment and diagnosis of heart disorders such as cardiac ischaemia,
CC heart failure and angina. This sequence represents the human RGS11
CC protein described in the disclosure of the invention
XX
SQ Sequence 31267 AA;
Query Match 9.4%; Score 3878.5; DB 6; Length 31267;
Best Local Similarity 21.2%; Pred. No. 4.6e-190;
Matches 1954; Conservative 1220; Mismatches 3544; Indels 2493; Gaps 317;
QY 10 PRFTTRKAF-VVSGVGDATLSQIIVGNPTFQVSWKQDQPVITAGARFLRAQDGLRLT 68
DB 6861 PKFVKLEASKIVKAGDSRSLCKIAGSPRIVVWFNEHELPA-----SDKYRMT 6911
QY 69 ILD-----LALGDSGVVCCARNAIGFAAAGLQVDAEACAEQAPHFLLRPTSI 119
DB 6912 FIDSVAVIQMNLSTESDGFCEAQNPAQSTSCSTKIV-----KEPVSFPPPIV 6964
QY 120 KVRGSEBATEFCRVGGSPRAVWSKGRIGEPDGRVRVVEELGEASALRIRARPRDG 179
DB 6965 ETLKNAEVSLECELSGTTPPFVWVKQRQL--RSSKKYKIASKNFHTS:HLNVDTSDI 7022
QY 180 GTVEVRAENPLGAASAAAL-----VVDSDAADTASRPGTSTAALLAHL 223
DB 7023 GEYHCKAQNEVSDTCVCTVKLKEPPRFVSKNLSLTVVAGEPBLQASIEGAQPIFQWL 7082
QY 224 QRRREAME-----AEGAPASPSTGTRTC----- 247
DB 7083 KEKEEVIRESENIRITFVENVATLQFAKAEANAGKYICQKNDGMRENMATLWLEPA 7142
QY 248 -----TVTEGKHARLSCTVYGEPEKPEVFWKDGQOLVTEGRRHVVEDAQENFVLKI 298
DB 7143 VIVEKAGPMVTYGETCTLECKVAGTPELSVWYKDGKLLTSQKH-KFSFYNKISSLRI 7201
QY 299 LFCQSDSLGTLCTASNLVGTYSVLVVREPAVP--FKKRLQDLEVRKESATFLCEV 356
DB 7202 LSVRODAGTTFYQNNVSKSCTAVVDSDRAVPPSFTRRLKNTGGVLGASCLCEKV 7261
QY 357 PQPS-TEAAMKBEITRLWASAKYIEEGTERRITVRNVSDADDVAVICETPE--GS--- 410
DB 7262 AGSPISVAMFHEKTIKIVSGAKYQTTFSQNVCTLQLNSLSDMDGNYTCVAANVAGSDEC 7321
QY 411 RTVAELAVQGNLLRLKPKRTAVRGDTAMF-CVELAVPVPVWLRNQEEVVAGRVAIS 469
DB 7322 RAVLTVOPPPFVKZ-PEPLEVLPGKNTFTSVIRGTPPFKVNFRGARELVKGDRCNIY 7380
QY 470 AEGTRHTLTIQCCLEDVQO----VAFMAGDCQSTR----- 502
DB 7381 FEDTVAELELEFNIDISQSGEVTVCVSNNAQASCTTRLFVKGLLKWIFFCFLYTDILVF 7440
QY 503 -FCVSAKPPLOP-----PVDPVKARMESSVLWSPPHGERPVTI---DG 547
DB 7441 FLCFLAVLCHLEPAAFKRLSDHSVEPKSIIILESTYT-----GTLPISVTWKKG 7492
QY 548 YLVEKKLGLTVTWIRCHAEAWATPELTADVABEENFQFRVSALNSFGQSPYLEFGTV 607
DB 7493 FNITTS-----KCNIVTTEKTCILEILNSTKEDAGQYSCIEINEAGRDVCGALVST- 7544
QY 608 HIAPLKAVRTPKVAQVVEGVTSPVLDLTVASAGE-----WFLDQALKASSVVEIHCDR 663
DB 7545 -LEPPYFV-TELEPLEAAVGSV--SLQCVAGTPEITVSWYKGDTKLRPTPEYRTFTN 7600
QY 664 TRHTLTIREVPASLHGAQLKFVANGI-----ESSIRMEVRAAPGLTANKPPAAAAREVLA 718
DB 7601 NVATLVFNKVN:NDSEYTCNAEN:IGTASSKTVEIOERQLPPSFARQ-----LKDIEQ 7655
QY 719 RLHFEAQLLAELSDQA-AAVWLKDGRTLSPGPKYEVQASAGRRLVLRDVARDDAGLYE 777
DB 7656 TVGLPVTLCRLNGSAPIQVCWYRDGVLRRDDENLQTSFVDNVATLKILQTDLSHSGQYS 7715

QY 778 CVSRGRITAYQLSVQGLARFLHK-----DMAGSCVDVAVAGGPAQRECETSEAH-VHVHWYK 832
DB 7716 CSASNPLGTASSABRLTAREPKSPFFDIKPVSDIVIAGESADFECHVTGQAPKMITWSK 7775
QY 833 DGMELHSGSERPQSDVGRHRLVAATVTRQDEGTYSR-----VGEDSVDFLRLRSEPKV 888
DB 7776 DNKEI-RPGNYTITCVGNTPELRLILKVGKSGGYTCOATNDVGDKMCQAQLSVKEPKK 7834
QY 889 VFAKQLARKLOAEAGASATLSCEVA-OAQTEVTWYKDGKLLSSSKVCMCATCCTRL 947
DB 7835 FVKLEASK--VAQGESIQLECKISGSPKLSVFNFRNDSHESWYNKNSFINSVALL 7891
QY 948 VVQOAGQADAGEYSCEA-----GGQRLSFHLDVKEPKVFAKQDVAAHSEVQAEAGANATLS 1003
DB 7892 TNEASAEDSDGIYCEAHNGVGDASCSTALTATVKAPEVTKPE-----SPVGLKGSVDILO 7947
QY 1004 CEVAQAQA-EVWYKDGKLLSSSKLVHVEAKCRRRLVVQAGKTADAGDYSCAERGQ--- 1059
DB 7948 CEISGTGPPPEVWVWDRKQVRNSKFKITSKHFDTSHLINLEASDVGEYHCKATNEVGS 8007
QY 1060 -RVSPRLHITPEKMFVFAKEQSVHNEVQAEAGASAMLSCEVAQAQ-TEVTWYKDGKLL--- 1114
DB 8008 DTCSGSKVKEPP-RPVKLS---DTSILGDVDELRAIVEGFQIPVWMLKDRGEVIRE 8063
QY 1115 SSSSKVGMVKGCTRLRLVLPQAGKADAGEYSC-----EAGQQRVSPHLHITBPKGVFAKEQ 1170
DB 8064 SENTRISFIDNIATLQGSPEA--SNSGKVIQIKNDAGMRECSAVLTVLEPARIEKPE 8121
QY 1171 SVHNEVQAEAGTAMLSCEV-AQPQTEVTWYKDGKLLSSSKVMEVKCCTRLRVVQVVG 1229
DB 8122 ----PMTVTGPNFFALECVTGTPELSAKWFKDGRLELSADSKHITFINKVASLKIPEAE 8177
QY 1230 KADAGEYSCE-----AGQQRVSPQLHITE---PKAVFAKEQLVHNEVTRTEAGASATLSCEV 1282
DB 8178 MSDKGLYSFEVKNVSKNCTVSVHVSRIVPSPFIRKLK---DVNAILGASVWLECRV 8233
QY 1283 A-QAQTEVTWYKDGKLLSSSK-----VRTEAG--C 1311
DB 8234 SGSAPISVGFODGNEIVSGPCQSSFSFENVCTLANLLEPSTDTGTYTCAANVAGSDEC 8293
QY 1312 MRQLVVQO-----AGQA----- 1323
DB 8294 SAVLTVOEPPSPQTPDSVEVLPKMSLTFTSVIRGTPPFKVKWFKSGRELVPGESCNISL 8353
QY 1324 -----DAGEYTC-----EAGQRLSFHLDVSEPKAVFAKEQLAHKRVQAEAA 1364
DB 8354 EDFVTELELFEVQPLESGDYSCLVNTDAGSASCTTHLFVKEP-ATFVK-RLADFSV--ET 8409
QY 1365 GATATLSCE-VAQAQTEVTWYKDGKLLSSSKVMEVAVGCTRLRVVQOACQADTGYSYC- 1422
DB 8410 GSPVILEATYTGTPPISVSWIKDEYLISSERCSITWTEKSTILEILESTIEDIYAQYSL 8469
QY 1423 ---EAGQRLSFSLDVAEPKWVFAKEQPVHREYQAAQAGASTTISCEV-AQAQTEVWYKX 1478
DB 8470 IENEAGQDICEALVSVLEPYPI---EPL-EHVEAVIGSPATLQCKVDGTPTETIRSMYKE 8525
QY 1479 GKLLSFSSKVRMEVAVGCTRLRVVQOAGQADAGEYSCEA---GSQRLSFHLHVAEPK--A 1532
DB 8526 HTKLRSAPAYPMQFKNNVASLVINKVDHSDVGEYSCKADNSVGNVAVASSAVLVTKARLPP 8585
QY 1533 VFAKEQASREVQAEAGTSATLSCEVAQAQ-TEVTWYKDGKLLSSSKVMEVAVGCTRL 1591
DB 8586 FPARK---LKDVTETLGFVAFECRINGSPELVQSWYKDGVLKDDANLQTSFVHNATL 8642
QY 1592 VVQOAGQADAGEYSCKA---GDQRLSFHLHVAEPK-VFAKEQPAHREYQAEAGASATL 1646
DB 8643 QILQTDQSHIGQYCNASNPGLGTASSAKLILSEHEVPPFFDLKPV--SVDLALGSGTGF 8700
QY 1647 SCV-AQAQTEVTWYKDGKLLSSSKVMEVAVGCTRLRVVQOAGQADAGEYSCE---AG 1701
DB 8701 KCHVTGTAPKITWAKNDREIRPGNYKMTLVNTATLTVLKVKGADAGQYTCVSNIA 8760
QY 1702 GORLSFELHVAE-----LEPQISRPCKRPELVVKEHBDIILTATLTPSAAVTWKL 1754

| | | | |
|----|-------|---|-------|
| Db | 8761 | KDSCSAQLGVQPPRFIKKLEPS-----RIVKQDEFTRYECKIGGSPKIVLYWK | 8810 |
| Qy | 1755 | DOVEIRRRKRHETASQDTHLTVHGAOVLDSAIYSCR-----VGAEGODFPVQVEEVAAK | 1810 |
| Db | 8811 | DETEIQSSKFRMSFVDSVAVLEHNLNVEDSGDYTCETHNAAGSASSSTSLKVKKEPPI- | 8869 |
| Qy | 1811 | FCRLLEPCVCGELGGVTTLACEL--SPACAEVVRGCGNTOPRVGRFQVVAEGPVRSLTVL | 1868 |
| Db | 8870 | FRKKPHPIETLKGADVHLECELOGTPP-FHVSWKDKRELSGKKIKIMSENFUTSHIL | 8928 |
| Qy | 1869 | GLRBADEGYCESRDDHSHSQAQLTVSV-----PRVKFMSGLSTVVAEBEGEATFOCVWS | 1923 |
| Db | 8929 | NVDAADIGEQCKATNDVGSDDTCVGSIALKAPPRFVKLSDISTVV--GKEVQLQTTIE | 8985 |
| Qy | 1924 | PSD-VAVVWFRD-GALLQSEKFAISQSGASHSLTISDLVLVEDAGQITVAEGASS-- | 1978 |
| Db | 8986 | GAEPISVWVFVKDGEIVRESNIIWISYSENATIQGRVBPASAGKYTCQIKNDAGMOEC | 9045 |
| Qy | 1979 | AALRVREAPVLFKKKLEPQTVBERSSVTEVELT-RPWPEL--RWTRNATALAPGNVE | 2034 |
| Db | 9046 | FATLSVLEPATIVEK--PESIKVTGTCTLECTVAGTPELSTKWFKDKGKELTSDNKKY | 9102 |
| Qy | 2035 | IHAEGARHLVHNVGFPADRFCCETPD-----DKTOAKLTVMRQV--RLVRGLQAVEA | 2088 |
| Db | 9103 | ISFNKVSGLKIIINAFSDSGVYFQVQNPVKGDSCTASLQVSDRTVPPSFTRKLKETNG | 9162 |
| Qy | 2089 | REQGTATMEVOL--SHADVDGSWTRDGLRFQOGPCTCHLAVRGPMTHTLTLGLRPEDSG-- | 2144 |
| Db | 9163 | LSGSSVMECKVYSGPPISVSMFHEGNEISSGRKYQITLTDNTCALTVNMLESDSDYT | 9222 |
| Qy | 2145 | LMVFKABGVHTSARLVVTELPVGSFRP-LQDVVTEKEKVTLECELSRPNVDVRLKOG | 2202 |
| Db | 9223 | CIA-TNMAGSDCSAPLTVREPPSFVQRPDPMDLVTGNTVFT-SIVKGTFFPSVSWFKGS | 9281 |
| Qy | 2203 | VELRAGTMAIAAOGACRSLTIXRCEPADQGVVYCDADAQASASVKVQGRTYTLVRRV | 2262 |
| Db | 9282 | SELVPGDRCNVLSDESVAELELFVDVT:SQSGEYTC-----I | 9317 |
| Qy | 2263 | LAEDAGIEQFVAENAEARQLRVKELPVTILVRPLRDKIANEKHRG-VLE-COVSRASAQV | 2320 |
| Db | 9318 | VSNEAGK-----ASCETHLYIK-APAKFVKRLND-YSIEKGKPLILEGTFTGTPPISV | 9368 |
| Qy | 2321 | RWFKGSQELQPGPKVELVDGLYRKLIISDVHAEDEDTYTC-----DAGDVKTSAQPFVEE | 2376 |
| Db | 9369 | TWKQNGINVTFSQRCNITTEKSAILEIPSTVEDAGQYNCYENASGKSDCSAQIILIE | 9428 |
| Qy | 2377 | QSI--TIVRGLQDVTVMEPAPAFECETSIPSVRPPKWLGLKTVLQAGNVG--LEQEG | 2431 |
| Db | 9429 | QKLPSPFSRQLRDVQETWGLPVVFDCAISGSEPIISVWYKDGKPLKDSPNVQFSPLDNTA | 9488 |
| Qy | 2432 | TVH-----RLMLRBTCTMTGPHFTVTKSRSARLVUS-----DIPVLTPLPEPK | 2478 |
| Db | 9489 | TLNIFKTRSLAGQYSCATNP-----IGSASSARLILITGKNPPPFDIRLA---EVDVAV | 9541 |
| Qy | 2479 | TGRELOSJWLSJCDPR-----PAPKAVOWYKDDTPLSPSEKFKMSLQGMALRILMPEA | 2533 |
| Db | 9542 | VGE-----SAUFECHVTGTQPIKVSNAKDSREIRSGGKYQISYLENSAHLTVLKVDKG | 9594 |
| Qy | 2534 | DAGVYRQOA-----GSAHSSTEVTVAREV--TVTGPLQDAEATEEGWASFSCE--LSHE | 2584 |
| Db | 9595 | DSGOYTCYAVNEVKDSC2AQLMIKERLIPPSFTKRLSETVBETEG-NSFKLEGRVAGSQ | 9653 |
| Qy | 2585 | DEEVEWLSNGMPLYNDYSHSHKGRHRTLVLKSIORA---DAGIVR-----ASSLKVS | 2635 |
| Db | 9654 | PIIVAWYKXNIEOPTSNCHITFK--NNTLVLO-VRKAGWMDAGLYTCVSNVNDAGSALCT | 9710 |
| Qy | 2636 | TSARLEVRVPVFLKALDLSAEERGTALQCEVSDPE-AHVVVRKDGVLQSPDKYDF | 2694 |
| Db | 9711 | SSIVKEPKPPFPDQHLTFTVTVSEGYVQLSCHVQSEPIRQIOMLKAGREIKPDRCSF | 9770 |
| Qy | 2695 | LHTAGTRGLVVHVDSPEDAGLYTCH---VGSBETRARVRVHDLHVGTIKRLKTHEVLEBG | 2750 |
| Db | 9771 | SPASGTAVLELADVAKADSGDYVCKASVAGSDTTKSKVTIKAEPIQTKRQIONVSEH | 9830 |
| Qy | 2751 | EGSCFECVLSHESADPPA/VWTVGGTKVSSSRFQATROGRKYILVVRBAAPSDAGEVVS | 2810 |
| Db | 9831 | QSATECEVSFDDAI--VTWYKGPTELTSQKYNPRNDRCHYMTIHNVTDPDEG--VYS | 9886 |
| Qy | 2811 | V-----RG-----LTSK-----ASLIVR----- | 2823 |
| Db | 9887 | VIARLEPRGEARSTAEIYLTTKKIELKLPDPIDPSRVPIPTMPPIRAVPEEIPPVVAPP | 9946 |
| Qy | 2824 | -----ER-----PAALIKP-----LEDQ | 2836 |
| Db | 9947 | IPLLPTPEKKPPKRIEESHERKVPKAPVEKKAPPKPKVIEKIEKTSRMEEE | 10006 |
| Qy | 2837 | WV-----AP----- | 2840 |
| Db | 10007 | KVQVTKVPEIKPAIPAPPEPKPEAGVPKTPSPSIEARRKLRPGSGGKPPDEAPFT | 10066 |
| Qy | 2841 | -----GEDVELRCELSRAGTPVHMLKDRKAIKRSQKYDW | 2875 |
| Db | 10067 | YOLKAVPLKFVKEIKDIILTESEFVGSASAIPECLVSPSTAITTWKOGSNIRESPKHPFI | 10126 |
| Qy | 2876 | CEGTWMLVIRGASLKDAGEYTCV-----EASKSTASLHVBEKANCFTTEELTNQVEBK | 2931 |
| Db | 10127 | ADGDKDKLHIIDVQLSDAGEYTCVLRNLGNKEKTSKALWBEELPVRFVKTLEEVTVVKG | 10186 |
| Qy | 2932 | TAVFTCKTEHPAATVTRK-GLLELRASGKHOPQEGTLTLTISALEKADSDTYTCDIG | 2990 |
| Db | 10187 | QPLYLSCELNKERDVWREKDKIVVEKPGRIVPGVIGLWRLALTINDADDDTAGTYTIVE | 10246 |
| Qy | 2991 | QAQ-----SRAQLLVGRRVHIIEDLEDVDVQEGSATFRCSIPANPEPVHWF--LDKTP | 3044 |
| Db | 10247 | NANNLESCSVKVVBEIRDVLWPKPIRDOHVKPKGTAFACDIA-KDTPNIKFKGYDEIP | 10305 |
| Qy | 3045 | LHANELNIDIAQPGYHVLTLRQLAKDSGTIYEAGDQASAAALRVTEKPSVSRDLTD | 3104 |
| Db | 10306 | AFPNDKTEI-URDGNHLYLKIKNAMPEDIAEYAVEIEGKRYPAKLTIGEREVELLKPIED | 10364 |
| Qy | 3105 | ATITEGEDTLVCETSTCDIPMCWTQDKTLRGSARQOLSHEGHRALQTLITGATLQDSGR | 3164 |
| Db | 10365 | VTIYEKESASPDABISEADI PQOMKLGELLRPSPTCEIKAEKGKFLTLHKVLDQAGE | 10424 |
| Qy | 3165 | YKCEAGGACSSIVRVHARPVRFQBEALKOLEVLEGGATLRCLVSSVAAPVKWCYGNVL | 3224 |
| Db | 10425 | VLYQALNAITTAITLVKBEILDFAVPLKDVTVPERROARFECVLTR-EANVWSKGPDI | 10483 |
| Qy | 3225 | RPQKYSIRQBGAMLELVVRNLRPDQSGYSCSFGDQTTSATLTVTALPAQFICKLRNKE | 3284 |
| Db | 10484 | KSSDKFDIADGKXHLVINDSQDDEGVYTAEEVGKTSARLFTVGTGRUKFMSPLEDQT | 10543 |
| Qy | 3285 | ATEGATATLRCELS-KTAPVWRKGSSETLRDGYCYLRQDGMCELOIRGLAMVDAEYS | 3343 |
| Db | 10544 | VKEGETATFVCELSHEKXHVVMFKNDKALHTSRVTLISSEKTHKLEMKVETLDDISQIK | 10603 |
| Qy | 3344 | CVCGEERTSASLITRMPAPHFIRLRHOESTEGATATLRCELSKAAVPEWRKGRESURDG | 3403 |
| Db | 10604 | AQVXELSTAOLKLEADPPTVTKLHDKTAVKDEITLKEVSKDVVPVKFKGEEIVPS | 10663 |
| Qy | 3404 | DRHSLRDGAVCELOICGLAVADAGEYSCVCGEERTSATITVLRALPAKFTFEGRLNEAVE | 3463 |
| Db | 10664 | PKYSIKADGLRILIKIADLKDKGEYVDCGDTKTKANVTVEARLIKVEKPLYGVFV | 10723 |
| Qy | 3464 | GATAMLWCELS-----KVAPVWRKGPENLRDGYLRLQEGTRCELOICGLAMA | 3513 |
| Db | 10724 | GETAHFIEISEPPVHGWKLKGQPLTASPDCEIIEGKKHIL-----ILHNQCLGM- | 10775 |
| Qy | 3514 | DAGEVLCVCGOERTSATITRALPARFIEDVYKQAEAGATVLCQELN--SAAPFVWRKG | 3572 |
| Db | 10776 | -TGEVSPQAANAKSAANLKVKEPLFIITPLSDVKVPEKDEAKFECEVSRPKTFRWLG | 10834 |
| Qy | 3573 | SETLRDGRYSLRDGTCKEQLQIRGLAMADTGEYSCVCGOERTSAMLTVRALPIKTEGL | 3632 |
| Db | 10835 | TCEITGDRELIKDGKTHSNVKSAAFEADKAFEDKGHGKGLIIEGILKFLTJPL | 10894 |

QY 3633 RNEBATEGATV----- 3644
: : : : :
Db 10895 KDVTAKESAVFTVELSHDNRVVKFKNDQRLHTRSVQMDBGKTHSITFKDLSDDT 10954
QY 3645 -----LRCELSM-APVEMWKGE 3662
: : : : :
Db 10955 SQIRVEAMGMSSEAKLTVEGDPYFTGKLQDYTGVEKDEVILQCEISAKADAPVKWFKDGK 11014
QY 3663 TLRDGRHSRLQDARGCELOIRGLVAEDAGBYLQWCGKERTSAMLTVRAMPSKFIIEGLRN 3722
Db 11015 EIKPSKNAVIRKADGKRMILILKALKSDIGQYTCGCTDKTSGLKDIEDREIKLVRPLHS 11074
QY 3723 EE----- 3724
Db 11075 VEVMETETARPETBEISDDIHANWKLGKGEALLQTPDCEIBEKGKHSILVHLNCRDLDTGG 11134
QY 3725 -----ATGDDTALWCELS-KAAPVEMWKGEHTL 3752
Db 11135 VDFQANVKSAAHURVKPRVTGLRLPLKDVTVTAGEATFOCELSYEDIPVEMWLKGGKL 11194
QY 3753 RDGRHSRLQDGRSCRELQIRGLVAVDAGEYSCVCGQERTSATLTVRALPARFIBDVKNQE 3812
Db 11195 BPSDKVVRSEGVKHTLRLDVKLEDADEVQLTAKDFKTHANLVKBPVPVEFTKPLEDQT 11254
QY 3813 ABEGTAVILOCELSKA-APVEMWKGESETLRGDRYSLRQDGTRELQIHGLSVADTGEYS 3871
Db 11255 VEEGATAVLECEVSRENAKVWFKNGTEILSKKYEIVADGRVKLVTHDCTPEDIKTYT 11314
QY 3872 CVCGQERTSATLTVRAOPVFPREPLQSLQAESEGTATILQCELSBPTATVVMKSGLOLQA 3931
Db 11315 CDKDFKTSNVLNVPVPHVEFLPLTDLQVREKEMARFECELSRENKVKWFKDGAIEKK 11374
QY 3932 NGRREPRLOGTAEVLVLODQREDTGEYTCGGSQATSATLTVTAAPVRFURELOHQEVD 3991
Db 11375 GKXYDIISKGAVRILVINKCLLDDAEYSCEVARTARTSGMLTVLEEAFTKNLANIEVS 11434
QY 3992 EGGTAHLCELSRAGASVEMWRKGSLOLPCKAYQMVODGAAEALLVRGVEODEAGDVTCD 4051
Db 11435 ETDTTLKVCESKPGAEVINYKGEDEIETGRYEILTEGRKILVQNAHLEDAGNYCR 11494
QY 4052 TGHQTSMASSLVRPRPKTKRLQSLQEOBTGDIARLCCOLSDAASGAVVQMLKGVELHA 4111
Db 11495 LPSRTDGVKVKHELAFAFISKPNLEILEGEKAEFVCSIS--KESFPVQWKRDDKTLES 11552
QY 4112 GKPYEMRSOGATRELLIHQLEAKDTGEYACTVGGOKTAASLRVTEPEVTIVRGLVDARVT 4171
Db 11553 GDKYDVIADGKRVLVKXDATLQDMGTYYVMVGAARAAHLTVIE-KURIVVPLKDRVK 11611
QY 4172 ADEDEYFSEVSRAGATGVQWCLQGLPLQSNTEVEVAVRDGRHTRLRKGVTPEDAGTVS 4231
Db 11612 EQQEVVFNCEVTEGAKAKWFEERAI FDSKY--IILQKDLVYTLIRDAHLDDQANYN 11669
QY 4232 FHLGNH-----ASSAQLTVRAPEVITILEPLQDVQLSEGDASFCQLRSRASQOEARWALG 4286
Db 11670 VSLTHRGENVKSAANLIVEEDRLRIVEPLKDIETMEKKSVTFWCKVNRNLN-VTLKWTKN 11728
QY 4287 GYPLQANENNDITVEGTLHLTLKVKLTLEDAGTVSFHVGTCSSEAQLK-VTAKNTVVRG 4345
Db 11729 GEVFPDRNVSRYVK-KYHMLTIDKCGFPDEGEIIVTAGQDKSVAELLITIEAPTEFVEH 11787
QY 4346 LENVEALEGEALFEQOLSQOPEVAATHLLDDDEPVRTSENABVFFENGLRHLLLNLR 4405
Db 11788 LEDQVTPEPDDAVFSQLSR-EKANVWYRNGREIK--EGKKYKFEKDGSIHRIIICDR 11844
QY 4406 PQDSCEVTVFLAGMVTSAFLTVGRWFLTEILEPLKNAVVRAGAQAARFTCTLSEAVPVGAS 4465
Db 11845 LDDECEYACGVEDRSARLFFVEEIPVEIIRPPQDILEAPGADVVFLELNLN--DKVEVQ 11902
QY 4466 WINGAAVQDDSDWTVTADGSHQALLRSQAQPHAGEVTFACRDVAASAR----- 4516
Db 11903 WLRNNMVVQGDKH-QMMSEGIHRLIQICDIPROQGEVRFYAKDKEARAKLEAAAPKI 11961
QY 4517 ----- 4516
Db 11962 KTADQDLVVDVGKPLTMVVPYDAYPKAEAEWFKENEPLSTKTIDTTABQTSFRILEAKG 12021
QY 4517 -----LTVLGLPDPPEDAEVAWSSHTVTLSSWAAWPMDSGGGGLC 4555
Db 12022 DKGRYKIVLQNKHGKAEAGFINLVLDVPGPVNLEVTETFDGEVSLAMEEPLTDGSGKII 12081
QY 4556 GYRVEVKEGATQWMLCHLHELVPGEVCDGLAPGET-YRFRVAAGVPVAGAGPHVLPQTV 4614
Db 12082 GYVVERDDIKRTWTLATDRAESCEFTVTGLOKGGVEYLFVRVSAARNVGTGPVETDNPV 12141
QY 4615 RLAEPKPKVPPQP-----SAPESRQVAAAGEDVLELEVAEAGEVHWHKM 4660
Db 12142 E-ARSKYVPGPPLNVTITDNRFGVSLTWEPEDVGDGAETINYVIELRDKTSIRWDTAM 12200
QY 4661 ERIQPGRFEVVSQGRQQLVTKGFTAEDQGBYHCOLAOGSICPAAATFOVALSPASVDE 4720
Db 12201 T-----VRAEDLSATVTDVVEGEYSFRVRAQNRIGVGK-----PSAATPFVKVADPIER 12251
QY 4721 APOPSLPPEAAQEGDLHLWEALAKRMRSRPTLDSISELPEEGRSQRLOPQAEVAP 4780
Db 12252 SPVNLITSDQTSQSSVOLKWEPLKD---GGSPILGYIIERCEE-GKONWIRCNM-KLVP 12306
QY 4781 DLSEGYSTADELARTGDADLSHTSSDDESRAGTSPSLVTLKKAGRFGTSPSLASKVGAPAA 4840
Db 12307 ELTYKVTGLEK---GNKYLVRVSAENKAGVSDPSEI----- 12339
QY 4841 PSVKPOQOQEPPLAAVRPPLGLDLSKTD-LGDPMSHDKAAVK-----TQAAEFKGYK 4887
Db 12340 -----LGPLTADDAFVEPTMDLSAFKDXGLDEVIPNITILVPTGY- 12380
QY 4888 VRKEMKQOEGPMPSHTFGTEAQVGDALRECVASKADVARWLRKDGVELTDGRHHHD 4947
Db 12381 -----PRPTATWCFGDKVLETGRVRKMTLSA----- 12407
QY 4948 QLQDGTCSLLIAGLRADAGCYTCQVSN-----RFGQVTHSA 4984
Db 12408 -----YAEVLISPSERSDXGIVTLKLENRVKITSIDVNIARPSAPKELKFGDITKDS 12462
QY 4985 CVTVSGSESAESSSGELDDAPRAARLRHLRFRK----- 5021
Db 12463 VHLT---WEPPDDDGSPITGVVKEKREVRKTKWMDVFDLFTVDPDLVQKEYLF 12518
QY 5022 -----SPAYSDEBELFLSADEGPAEPEPADWQTYREDEHFI----- 5058
Db 12519 KVCARNKCGPGEPAVY-DEPVNMSTPATVPDPENVKRDRTANSIFLTWDPKPKNGGSR 12577
QY 5059 -----CIR-----FEALTARQAVTRFCE--MEATLIGIGVEIKLVEQGPVRVEM 5100
Db 12578 IKGYIVERCPRGSKWVACGEPVAETKMEVTGLEGKWA-----YRVKA 12622
QY 5101 CISKETAPVVPPEPLPSLLTSDAAPVFLTE--LQNOEVQDGYPSFDCVVTQGPMPSVR 5158
Db 12623 LNRQGASKPSRPTTEECAVDTOEAPFLFDVLLAGLTVKAGTKIELPAITVTKPBPKIT 12682
QY 5159 WFKDGLLEBDDHYMINEDQGGHQHIIITAVPADMGVTRCLAENSMGVSSSTKAEILRV-- 5216
Db 12683 WTKADMILKQDKRITI-ENVPKASTVTIVDSKESDGTGYIIIEAVNCGRATAVVEVNVLD 12741
QY 5217 -----DLTS-----TDYDTAADATESSSYFSAQYGLSSREQ 5247
Db 12742 KPQPPAAFDITDVTNESCLLTWNPPRDDGSKLTINYVERATDSEVHWKLSSTVKDNTF 12801
QY 5248 EGTESTTDEGQLPQWVEEL-----RDLQVAPGTRLAKFOLKVKGYK-----APR 5291
Db 12802 KATKLIENKEYIFRVAENNMVGEVPQASPIIT--AKYQDPDPGPPTRLEPSPDITKDAVT 12859
QY 5292 LYWFK-----DQOPLTASAHIRM---TGKKIL-----HTLEIISVTREDSGQYAAIS 5336
Db 12860 LTVCEPDDDGSPITGWWVERLDDTDKVRCKNMFVKOTTYRVKGLTNKKKRVFRVLA 12919
QY 5337 NAMG---AAYSARLLVRGDEPEEKPSADVHEQLVPPRML-----RFTPKV 5382

| | | | | | | | |
|----|-------|--|-------|----|-------|---|-------|
| Db | 12920 | NLAGPKPSKSTEPILIKDIPDPS-----PPRWLEVINITKNTADLKWTVPK | 12968 | Db | 13705 | POSEMIWVTSTLRHCKYSVTKLIEGKEYLFRVRAENRFGPGPPCVSKPLVAKDPFGPPDA | 13764 |
| Qy | 5383 | KKGSSITTSVKVEGRPVPTVW-----LREEAERGVLKGP | 5418 | Qy | 6258 | P-----RGAPAL-----QETG----- | 6268 |
| Db | 12969 | DGGSPTNVI-VEKDVRRKGMQVDTTVKOTKCTVPTLTGSLVYFVRAENA-----IGQ | 13024 | Db | 13765 | PKPIVEDVTSNMLVKMNEPKDNGSPILGYWLEKREVNSTHWSRVNKSLSNALKANVDG | 13824 |
| Qy | 5419 | D-----TPGYTVASSAQOHSILVLDVGRHQGTGYTTCIASNAQAALCSA | 5462 | Qy | 6269 | -----SQPPVTGTSAPAVPRVPQPLLHHEGPEOEPEAIARAQ | 6306 |
| Db | 13025 | SDYTEIEDSVLAKDTFTTTPG-----PPYALAVVDVTKRHVDLKWBPXKDGGRPI--- | 13074 | Db | 13825 | LLEGLTVYFVCAENAAAGPKFSPSPDKTAHDPISPFGPPIPRVTDTSSITIEL---- | 13879 |
| Qy | 5463 | SLHVSGLPKVBEQE-----KVKEALISTFLOGTTOAISAOGL----- | 5499 | Qy | 6307 | EWTPVPRMEGAWPG-----AGTCELLWD-----VHSHVVRETO-RTYTYQAIQ | 6350 |
| Db | 13075 | QRYV-----IEKRELGRWVKAGTAGDCNFRVTDVIEGTVEQFQVRAENEAGVGHPS | 13129 | Db | 13880 | EWEPFAFNGGGEIVGYFVDKQLVGTNE--WSRCTERMIKVRYTVKEIREGADYKLRVSA | 13937 |
| Qy | 5500 | -----ETASPAD-----LGGQRKEEPLAAK--EALG-----HLSLAEVGTGE | 5534 | Qy | 6351 | THTA-----RPPSMQVITIE--DVQAQTGGTAQFEALIEGDPQPSVTM | 6390 |
| Db | 13130 | EPTIELSTEDPTSPSPPLDLHVTDAGRKHTAIANKPEKNGGSPILGYHVEMCPVTEK | 13189 | Db | 13938 | VNAAGEGPPGETQPVTVAEPOEPFAVELDVSVKGGIQTIMAGKTLRIPAVVTGRPVPTKVW | 13997 |
| Qy | 5535 | FLO-----KLTSQITEMVYSAKITQAK----- | 5555 | Qy | 6391 | YKDSVOLVDSTLSQOQBGTTYSVLVRHVASKQAGVYTCLAQNTGGQVLCKAELLVLGD | 6450 |
| Db | 13190 | WNRVNSRPIKOLKFKVEGVVDPKEYYLVRVAVNAIGVSEPEISENVVAKDPCKPTID | 13249 | Db | 13998 | TKEEGEL-DKDRVWIDNVGTKSELIKDALRKDHGRYVITATNSCGSKFAAARVEVF--- | 14053 |
| Qy | 5556 | -----LQVPGGDSDEDSKTPSASPRHGRSRPSSIOESSSE | 5593 | Qy | 6451 | NEP-----DSEKQSHRRK--LHSFYEVKEIEIGVGFVKRVQHKGNKILCAAKFTPL-RS | 6503 |
| Db | 13250 | LETHDIIIEGKLSIPVPRFRAVPVTVSHWKGKEVKASDRLTMKNDHISAHLEVPKSV | 13309 | Db | 14054 | DVPGPVLDLKPVVTVNRKMCLLNWSDEDDGSEITGFI--IERKDAKVHTWRQPIETERS | 14111 |
| Qy | 5594 | DGDARGETFDIYVVTADYPLGABQDAITLREGQYVEYLDAAHPLRWLVRTKTPKSSPSR | 5653 | Qy | 6504 | R-----TRAQAYRER-----DILAA--LSHPLVTGLLDOFE-TRKTLI | 6538 |
| Db | 13310 | RADA-----GIYTTILEN-KLGSATASIN-----VKVIGLPGCKDIKASDITKSS--- | 13354 | Db | 14112 | KCDITGLLEGQYKERVIAKNGKFGCPPEIPIILLAVDPLGPTSPESLTVTERTKSIIT | 14171 |
| Qy | 5654 | QGWSPAYLDRRLKLSPEWGAABEPPEGEAVSEDEYKARLSVSIQELLSEQAFVEELQ | 5713 | Qy | 6539 | L-----TILE-----LCSSELLDLRLKGVVTEAEVKVY | 6567 |
| Db | 13355 | -----CKLT--W-----EPPEFDGG-----TPILHYVLERRA----- | 13380 | Db | 14172 | LWKESPRNGSGSPIQGYIIEKXRDHDKDPFERVNRKLCPTTSFLVENLDEHOMYFRVKA- | 14230 |
| Qy | 5714 | FLOSHHLOHLRCPHVPFIAGQKAVIFRNVRDIDGRHSSFLQBLQOCDTDDVAMCFIK | 5773 | Qy | 6568 | IOQLVEGLHYHSHGVHLHDIKPNSILMVHPAREDIKICDFQAQNTIPABELOFSQYQSP | 6627 |
| Db | 13381 | -----GRRTYIPV-MSGENKLSW-TVKDILPNGEYFFR-----VK | 13413 | Db | 14231 | VNEIGESPSPLNVVIOQDDEVPTIKLRLSVRGDTIKVKAGEPVHI-PADVTGLPMPKI | 14289 |
| Qy | 5774 | --NOAAFEQYLEFLVGRVQABSVVSTAIQEFYKYABEALLAGDPSQPPPPLOQHVLEQ | 5831 | Qy | 6628 | EVVSFEIIOQNPVSPASDIWANGVI-SYLSLTCSPFPAGESORATLLNLVLEGR----- | 6679 |
| Db | 13414 | AVNKVGGEYIEL-----KNFVIAQDPKQPPDPVDEVEHN | 13449 | Db | 14290 | EWKNETVIEKP-TDALQITKEEVSRSSEAKTELSIPKAVREDKGTVTYTASNRLSGVFRN | 14348 |
| Qy | 5832 | PVERVQRYQALLKELIRNKAARNQCALLEQAVAVSALFORAENKLHVSIMENYPTGLE | 5891 | Qy | 6680 | -----VSWSSPMAAHLSEDAKDFIKATLQAPQAPSA | 6713 |
| Db | 13450 | PT-----AEAMTITWKPL-----YDGGSK | 13469 | Db | 14349 | VHVEYVDRPSPPRNLAVTDIKAESCYLTDAP-----LDNGSEITHVVIDRDRSRKA- | 14403 |
| Qy | 5892 | ALGEPHQGHIVVEGAPGAMPKGNHNVFLFRNHLVTCXPRRSRTDTVYVFRNMW | 5951 | Qy | 6714 | QCLSHWFLKSNPABEAHF-----INTKQLKFLLAGRNRWQSLMSY-----KSILVNR | 6762 |
| Db | 13470 | IMG-----YIIEKIAKEER-WKRCNEHL-----VPILTATKGL- | 13503 | Db | 14404 | -----EWBEVNTATVEKRYGIWKLIPINGQYEFVR-----RANKYGISDECKSKVVIQ | 14452 |
| Qy | 5952 | KLSSIDLNDQVEGDDRAFEVWQEREDSVKRYLLQARTAIKSSWKEICIGIQORLALPVW | 6011 | Qy | 6763 | IPELLRGPPDPSLGVARHLCRDTCGSS--SSSSSSSNELAPFA-----RAKSLPP-SP | 6813 |
| Db | 13504 | -----ZEGKEYQFVRAENAAGIS-----EPSRATPPT | 13531 | Db | 14453 | DPYRLPGPGPKV-LAR-----TKGMLVSWTLPDNGGSPITGYWLEKREESPTWSR | 14506 |
| Qy | 6012 | R-----PPFEELADCTAEL--GETVKLACRVGTGPKPVSIVYKDGKAVQVDPHHILIEDP | 6066 | Qy | 6814 | VTHSPL-----LHPRGFLRPSASLPEEAESERSTEAP | 6846 |
| Db | 13532 | KAVDFIDAPKVLRTSLVLRGDEIALDASIGSPYTIWIKDENVI----- | 13579 | Db | 14507 | VSRAPIITKVLKGVFEFNVPRLEGVKYQFRAMANAAGIGPSE--PSDPEVAGDPIFP | 14564 |
| Qy | 6067 | DGSCALILDSITGVDSDQYMCFAASAAGNSTGLGKILVQVPPFVNVKVRASPRVEGEDAQ | 6126 | Qy | 6847 | APPASPE-----GAGPRAAQCVPHSVIRSLFVHQAGSEPHGALAFGRSRHP | 6895 |
| Db | 13580 | -----VPEBI--KQGAAPLVRRKGE | 13598 | Db | 14565 | GPSPCEVQDKTKSISLGNKPPKADGGSPITKGYI-----VEMOEBGTTDKWRVNEP | 14616 |
| Qy | 6127 | FTCTIEGAPYQIRWYKDGALLTGKNTKFTLSBFRSGLLVLVIRAASKEDGLGYECELVN | 6186 | Qy | 6896 | ARRRHLLKGGYIAGALPGLREPLMEHRV--LEBEAAREEQATLLAKAPSPETALRLPAS | 6952 |
| Db | 13599 | VQ-----EEEPF-----VLPLTQRL-SIDNSKKGESQLRVRDLSRPHGLYMKIVEN | 13644 | Db | 14617 | DKLITTCB-----CVVFNLKE-LRKYFRVKAVNEAGESE-----PS-DITGEIPAT | 14661 |
| Qy | 6187 | RLQSARASABLRITOS-----PMLQAOEQCHREQLVAVE-----DTILERADQB- | 6230 | Qy | 6953 | GTHLAP-----GHSLSLEHDSPTSPSPSEAC--GEAQLRPSAPSGAIFRDMGHFQGS | 7004 |
| Db | 13645 | DHGIAKAPCTVSLVDTGPGPINFVFDIRKTSVLCKWEPPDLDDGGSEIINTYLEKDKTK | 13704 | Db | 14662 | DIQBEPEVFDIGAQDCL-----VCKAGSQIRIPAVIKO----- | 14695 |
| Qy | 6231 | -----VTSVLK-----RLLGPK-----APGPS-----TGDLTGPGPC | 6257 | Qy | 7005 | KQLPSTGCHPGHTAOPERPSPDPSPMGQPPAPFCHPKQSGAPQEGGSPHAPVAPCPGSPPG | 7064 |
| | | | | Db | 14696 | -----RPTPKSSW----- | 14703 |

QY 7055 SCKEAPLVSPFFILGQOAPAPAKASPLDSKMGPGDISLP--GRPKPGCCSPGASQ 7122
Db 14704 -----EFDGK-----AKKAMKLETAENSSVILPECKRSHGKYSI--TAKN 14743
QY 7123 ASSSQVSSLRVSSQVGTGPGSLDAEGWTOEADLSDSTTLQRPQEQATMRFSLGGR 7182
Db 14744 KAGQKTANCRV---KVMVDPGPPKDLK-----VSDIT-----R 14773
QY 7183 GGYAGVAGYGTAFAGDAGGMLGQGPMMARIWAVSQSEEEBEQBEARAESEBQBEARA 7242
Db 14774 GS-----CRLSKMPDDGGDRIKGVY----- 14795
QY 7243 ESPLPQVNSARPVEVGRAPTRSSPEPTWEDIGQVSLVQIRDLSDGAB----- 7290
Db 14796 -----IBKRTID--GKAWTKNP-----DCGSTTFV--VPDLLSEQQYFRVRAENRFG 14840
QY 7291 ---AADTISLSDISEVDPAYNLSDLYDKYLPFFFMIPRKVPKSAQPEPPSPMAEEELAE 7347
Db 14841 IGPPVETIQTAR--DPIY-----PPDPPIKLIKLG 14870
QY 7348 FPEPT-----MPWPGELPHAGLEITESEEDVDALLAEAAVGRKWKSPSRSLPHFPGRH 7403
Db 14871 ITKNTVHLNWKPKKNDG--GSPVTHYI--VECLAWDPTGTKKEAW----- 14911
QY 7404 LPLDPAELGLRERVKASVEHLSRLKGRPGLEKEGPPRKKPGLASFRLSGLK----SW 7459
Db 14912 -----RCKNRDVELOFTVDEGEVE-----FRVAVNAAGVSK 14949
QY 7460 DRAPTFLEBELDET-----VVLGOSVTLACQVSAQPAQAATWSXDGAPLSSS 7507
Db 14950 PSATVGPVTXKDOTCPSIDLKEFEVEEGTDVNIKIVKIGVPPPTLTWFKAPPKPDNK 15009
QY 7508 RVLISATLKNQOLL--TILVVAAE-----DLGVYTCSVSNALGTVTTGVLEKARPPSSP 7561
Db 15010 BPVLVDTHVNLVWDDTCTLVIPOSRRSDTGLYITIAVNLGTASKENLNLVLRPG--- 15066
QY 7562 CPDIG-----EYVADGVLLWKPVESYGP--VTYIVO--CSLEGGSMTTLASDIFDCCYL 7612
Db 15067 -PPVGPPIFESVADQNTLSWFFPKDGGSKITNVVIEKREANKRTWVHVSSEPKECTYT 15125
QY 7613 TSKLSRGTYTFRACVSKAGWG-PYSSPSE 7642
Db 15126 IPKLEHGYVFRIMAONKYGIGEPDSEPE 15156

RESULT 15
AAU05396

AC AAU05396 standard; protein; 26926 AA.
XX
DT AAU05396;
XX
DT 24-OCT-2001. (first entry)
XX
DE Human titin (connectin) protein sequence.
XX
KW Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
KW titin-related disease; zebrafish; heart failure; heart disease.
XX
OS Homo sapiens.
XX
PN WO200151666-A1.
XX
PD 19-JUL-2001.
XX
PF 12-JAN-2001; 2001WO-US001212.
PR 12-JAN-2000; 2000US-0175787P.
XX
PA (GEO) GEN HOSPITAL CORP.
XX
PI Fishman MC;
XX

DR WPI; 2001-451869/48.
DR N-PSDB; AA050390.
XX
PT Determining if a subject has or is at risk of developing a titin-related
PT disease or condition, particularly heart failures, comprises detecting
PT the presence of a mutation in the titin gene.
XX
PS Disclosure; Page 57-111; 114pp; English.
XX

CC The present sequence representing human titin (also known as connectin)
CC is described in an invention relating to a novel method for determining
CC whether a subject has or is at risk of developing a titin-related disease
CC or condition. The method comprises analysing a nucleic acid of a sample
CC from the subject and detecting the presence of a mutation (e.g. the
CC pickwick mutation in the cardiac specific exon N2B) in the titin gene,
CC which indicates that the subject has or is at risk of developing a titin-
CC related disease. The zebrafish which has a phenotype similar to mammalian
CC heart failure is used as a model. The method is useful for detecting an
CC increased likelihood of heart disease, such as heart failure, in a
CC patient, so that appropriate intervention can be instituted before any
CC symptoms occur. The method may also be used to facilitate determination
CC of etiology of an existing heart condition, such as heart failure, to
CC identify compounds that can be used to treat or prevent heart conditions,
CC in prenatal genetic screening, e.g. to identify parents who may be
CC carriers of a recessive titin mutation. Compounds identified using the
CC methods may be used to treat patients that have or are at risk of
CC developing heart disease, e.g. heart failure
XX

Sequence 26926 AA;

Query Match 8.7%; Score 3580.5; DB 4; Length 26926;
Best Local Similarity 20.2%; Pred. No. 1.1e-174;
Matches 1930; Conservative 1288; Mismatches 3661; Indels 2663; Gaps 320;

QY 4 PQSGAPRELTRPKAFVSVGKDATLSQIVGNFTPQVSWKXQDQVTVAGARFLA---Q 60
Db 1030 PGEPAAPYFITKVEVQKLVGEGSVVFGVCGVGNKPHVYWKSGVPLTTGYRYKYSYKQ 1089
QY 61 DGLYLRLTLDLALGSGQVVCRAINAIGEAFAAVGL-----QVD 100
Db 1090 TGEC-KLVISMTFADAGEYTIWRNKHGETSASLSLEADYELLKMQEENLYQTQT 1148
QY 101 A---EACACAPQAPHL----- 113
Db 1149 AFVQPEVEGTAPGFVYSEYEKEQALIRKMAKTVVVRYTVVEDQEFHISFEEL 1208
QY 114 -----LRPTSIRVREGSEATPCRV 133
Db 1209 IKEIYRIKTTLEELLEDGEBKMAVDISEAVESGDLRIKNVRIIEGKGVTFHCM 1268
QY 134 GGSFPAVSNKXDRRLGEPDGRVVRVBLGEASA-LRIRAAARPDGGTYEVRAENPLGA 192
Db 1269 SGYPLPKIAWYKDKRI--KHGERYQWDFLQDGRASLRIPVVLPEDEGIYTFASNIKN 1326
QY 193 ASAAAALVYDSDAADTASR--PCTSTAALLAHLQR-----REARABEGASAPP----- 240
Db 1327 AICSGKLYVEAPAPLGAFTYVITPLEVPSRIRSLSPRSVSRSPTRMSPARMSPARMSPARM 1386
QY 241 -----STGTR-----TCTVTEGKHARLSCYVTGEPKPTVWK 273
Db 1387 SPARMSPGRLEETDESQLERLYKPVFLKPVSKFCLGEGTARFDLKVWRPMPETFWFH 1446
QY 274 DQQLVTEGRH--VVVEDAQENFVLKLFCKSDRLGYTCTASNLVQYYSVLLVY--- 328
Db 1447 DGOQIVNDYTHKVIKEDGTQSLI--IVPATPSDSGEWTVWQNRAGRSISVILTVEAV 1504
QY 329 ----- 328
Db 1505 EHQVKPMFVEKLVKVNKIBGSRLEMKVRATGNPNPDIVLKNSDIIVPHKYPKIRIEGTK 1564
QY 329 -----R 329
Db 1565 GEAAALKIDSTVSQDSAWYTATINKAGRDTTRCKVNVEVEFAPPEPERKLIIPRGTYRAK 1624

QY 330 EPAVP-----FKKLODLEVREKESATFLCE---VP 357
Db 1625 EIAAPELEPLHLRYQOEOWEGDLVDKEQOKPFFKKLTSLRKRFGPAHFECRLTPIS 1684
QY 358 OPSTEAAWFKBETRL-----WASAKYGI-----EEGETERRLT 390
Db 1685 DPTMVVWLHDGKPLEAANRLRMINEFCYCSLDYGVAYSRDSGIITCRATNKYGTDTSA 1744
QY 391 VRNYSADDDAYVICTPEGR-----TVAEALVQCNL-----LEKLPRKT 430
Db 1745 TLIVKDEKSLVEESOLPEGRGLORIIELEMAHEGALTGVTTDOKEQKQPDIVLYPEPV 1804
QY 431 AVRVDGTAMF-CVELAVPVGPHVHRLNOEEVAVGRVAISAEGRTHLTITISQCCLEDVQ 489
Db 1805 RVLEGETARFCRVGTGYPQPKVMYLNQGLIRKSRFRVYDGI-HYLDIVDCSKSYDTGE 1863
QY 490 VAFVAG-----DCOTSTRFCVSAAPRKPPLQP-----PVD 518
Db 1864 VKVTAENPEGVIEHKVKLEIQOEDFRSVLRAPEPPEFHVHEPGKLOFEVQKVDREVD 1923
QY 519 -----PVVKARWESSVILSWPPHGE-----RPVTIDGY-----LVEKKKLGITYTW 560
Db 1924 TTEKEVVKRAERIHEKVPESBELRSFKRTEEGYEAITAVELSRKKODESYBE 1983
QY 561 I-----RCHAEBWATPELTVAD-----VABEGNFQFRVSALNSFGQSPYLEPFGTVHLAPKL 613
Db 1984 LIRKTKDELHWH--TKELTEBEKKALAEKGITITPTFKPKIELSPSME-----APKI 2034
QY 614 AVRPLKAVQAVEGEVTFSDLVASAG-----EMFLDQALKASS-VYEHICORTRH 666
Db 2035 FERQSQVIG--QSDAHFRVRV-----VGKPDPECEWKYKGVKIERSDRIYTWPDNVC 2088
QY 667 TLITREVPASLHGAKLXF-----VANGIESSIRMEVRAAGLUTANKPPAAAAAREVLARLHE 722
Db 2089 ELVIRDVTAE--DSASIMVKAINIAGETSSHAFLLVQAKLITFTQ-----ELQDVVAKEKD 2143
QY 723 E-AQLLAELSQAANAATWLDGRILSPGPKYEVQASAGRVLVVRDVARDDAGLYECV-- 779
Db 2144 TMAFTECSTSPFFVKVKYKDGVEHGDYKMHSDKXVHFLSLITLITDSDAEDYSCVLV 2203
QY 780 -SRGRIAYQLSVQGLARFLHKDMAGSCVDAVAGGPAQFECETSEAHVHVHWTYKDGMELG 838
Db 2204 EDENVKTTAKLIVSGAVVEFKELQD--IEVPESYSGELECIYSPENIEGKYHNDVELK 2261
QY 839 HSGERFLOEDVGTTHRLVAATVTRQDEGTYSRCVGEDSVDFRLRVSEPKVVFVFAKEOLAR 898
Db 2262 SNGKYTTITSRRG-RQNLTVKDVTKEDQGEYSFVIDGKKTKCKLM-KPRPIALQGLSDQ 2319
QY 899 KLOABAGASATLSCEVAQAQTEVTVYKDGKLSKSSKVCMEATGCTTRRLVVOQAGQADAG 958
Db 2320 KVC--EGDIVOLEVKVSVLESVBGMKDGQEVQPSDRVHIVDKQSHMLLIEDYTKEDAG 2377
QY 959 EYS-----CEAGQRLSHFLDVKPEKVPVFAKDQVAHSEVQABAGANATLSCEVAQAQ- 1010
Db 2378 NYSFTIPALGLSTGRVSVSDVITPL-----KDVNVIETGKAVLECKVSPDV 2427
QY 1011 AEVVMYKDGKLSLKVHVEAKCRRLRVVOQAGTKDAGYSCAARGQVSPRLHITER 1070
Db 2428 TSVKXVYNDEQIKPDDRQALVKGTKORLVINRTHASDEGPKYKLVGRVETNCLSVSEKI 2487
QY 1071 KMMFAKESQVNEVQABAGASAMLSCEVAQAQTEVTVYKDGKLSKSSKVMGVKCTTR 1130
Db 2488 KIIRGLRDLTCTEQ-----NVPEVELSHSGIDVLNFKDXKIKPSSKYKIEBAHKIKYK 2542
QY 1131 LVLPQAKADAGEYSCAAGQVSPFLHITPEKGVFAKESQVNEVQABAGATTAMLSCEV 1190
Db 2543 LTVLNMKDDBGKTYTFVAGENMTSGKL--TVAGATISK--PLTDQTVAES-QBAVFECEV 2597
QY 1191 AQPOTETVYKDGKLSKSSKVMGVKCTTRRLVVOQAGADAGEYSCAAGQVSPFLH 1250
Db 2598 ANPDSKGEWLDRDGKHLPLTNIRSESGHKRRLLIATKLDLDDIGEVTKYVATSKTSAK 2657

QY 1251 ITEPKAVFAKEOLVNEVTRTEAGASATLSCEVAQAQTEB-VTVYKDGKLSKSSSKVRIEAA 1309
Db 2658 V---FAVKIKKTL-KNLTVTET-QDAVFTVETHPNVGVQWIKNGVVLENEKVAISVK 2712
QY 1310 GCMROLVVOQAGQADAGEYTCAG--GORLSFHLN-----VSEPKAVFAKEOLAHKVKQAE 1363
Db 2713 GTIYSLIRKNCALVDES VYFRLGRLGASARLHVETVKI IKKPKDVTALN----- 2763
QY 1364 AGAIAITLSCEVAQAQ--TEVTYKDGKLSKSSSKVMEAVGCTTRRLVVOQACQADTGEYS 1421
Db 2764 ----ATVAFESVSHDTPVVKMFHKSVEIKPSDRHLVSEKVKHKLMLQNTLSPSDAGEYT 2819
QY 1422 CEAGQRLSFLSDVAEPKVVFAKEQFVH-----REVQAGASATLSCEVAQAQTEVVMY 1476
Db 2820 AVVGQLECKAKLFV-----ETLHITKMKNIETPETKTASFCEVSHFNVPSMWL 2869
QY 1477 KDGKLSKSSKVMMEAVGCTTRRLVVOQAGADAGEYSCAAGSQRSLSFHLHVAEPKAVPAK 1536
Db 2870 KNGVSIEMSEKFIYVQKGLHQLIIMTSTEDSABYTFVCGNDQVSATLTV-PIMITS- 2927
QY 1537 EQPASREVOAEAGTSATLSCEVAQAQTEVTVYKDGKLSKSSSKVMEAVGCTTRRLVWCEA 1596
Db 2928 ---MLKDINAEEKDITFEVTYVYEGISYKMLKNGVEIKSTDKQMTTKLTHSLNIRV 2984
QY 1597 GOADAGEYSCAGDORLSFHLHVAEPKVVFAKEQPAHREVOABAGASATLSCEVAQAQTE 1656
Db 2985 HFGDAADYTFVAGKATSTATLYVEARHIEFRKH---INKDIKVEKKRAMFECEVSEPDIT 3041
QY 1657 VTVYKDGKLSKSSSKVMEAVGCTTRRLVVOQAGADAGEYSCAAGQRLSFLHVAEILEP 1716
Db 3042 VQMKDDDELQITDRIKQKEKYVHRLIIPSTRSDAGKYTVAGGVNSTAKLVFEGEDV 3101
QY 1717 QISERPCRRPLVKEHEDIILTATLTPSAATVTLKDGVEI--RRSKRHETASQGDTH 1774
Db 3102 RI--RSIKKEQVIEKQRAVV--BFEVNEDDVDAAHWKDGIEINFQVERKHYVVERIH 3157
QY 1775 TLTVHGAQVLDSAIYSCVGAEGQDPFVQVE-EVAAKFCRLLEPVCGLGCTVTLACELS 1833
Db 3158 RMFISETROSAGEYTFVAGNRSSVTLYVNAPEPPQVLQELQVTVQSG-----K 3208
QY 1834 PA--CAEYVWRGNTQPRV-----GKRTQMAEGEVRSLTVLGLAEADAGEYVC 1880
Db 3209 PARFCAMI--SGRPQPKI SWYKEBQLLSTGFKCKFLHDGQEYTLILLIEAFPEDAAVTC 3265
QY 1881 ESRDHH-----TSAQLTVSVPRVK-----FMSGLSTVVAEGBEGEATFCQVSPS 1925
Db 3266 EAKNDYGVATTSSLSLSEVPEVSPDQBPYPPAIITPLOTVTSEGPAPFCRVSQSGT 3325
QY 1926 DVAVVWFRDGLLOPSEKFAISQSGASHLTSISLVLEDAGQITVEAGA---SSSAAL 1981
Db 3326 DLKVSWSYKDKKI KPSRFRMTQPEDITYLEI ABAYPEDEGTYTFVANNAVGVQVSSSTANL 3385
QY 1982 RVREAPVLFKKKLBPQ-----TVBERSSVTLEVLTR----- 2013
Db 3386 SLEAPESILHERIEQIEEMENKEFSSSFLSAEBSGLHSAELQLSKINETLELLESVPYP 3445
QY 2014 -----PWPELRWTRNATALAPKQVVEI----- 2035
Db 3446 TKFDSEKSGTGPIFIKEVSNADISMGDVATLSVTVIGIPKXIQWFFNGVLLTTSADYKP 3505
QY 2036 HAEGARHRLVHNVGFDADRGFFGCETPDD-----KT 2066
Db 3506 VFDGDHSLIILFTKLEDEGEYTCWASNDYKGTICSAYLKINSKGEHKGKDTETESAVAKS 3565
QY 2067 QAKL-----TVEMROVLVRGLQAV-EAREOGTATMEVQLSHADVDSGTEDGRFRQ 2117
Db 3566 LEKGGPCPPHFLKELKPIRCAQGLDPAIFEYTVVGBEPAVTV-----TWFKENKQL- 3615
QY 2118 QGPTC-----HLAVRGP--MHTLTLSGLRPEDSGLMWFAEGV-----HTSARLVV- 2161
Db 3616 ---CTSYYTIIHNPGSGTFFVNDPOREDSGLYICKAENMLGESCACAAELLVLEDDTD 3671
QY 2162 -----TELPVPSFR-PL-----QDVVTTEKEKVTLECELSRENVDVRV 2198

Db 3672 MTDPCKAKSTFAPEDPPOPLKGPAPVADLSQEIATFVKOTILKAALITEENQOOLS 3731
QY 2199 -----LKGVELRAGKTMAIAAQ-----ACRSLTIY----- 2225
Db 3732 EHIKANELSSQLPLGAQELQSIQDKLTPESTREFLCINGSIHFOPLKEPSSNLQOI 3791
QY 2226 ---RCEPADQGVVYCDADAQS----- 2244
Db 3792 VOSQKTFSGEGLMPEEPETOAVLSDTEKIFPSAMSIEQINSLTVEPLKTLAABEGNYP 3851
QY 2245 SASVQOGRTY-----TLIYRRVLAE----- 2265
Db 3852 QSSIEPPMHSLYVAEBVLSEKTVSDTNREQVTLQKQEAQASALLISQSLAEGHVES 3911
QY 2266 -----DAGEIQVA-----ENAESRAQLRVKE----- 2287
Db 3912 LQSPDVMISQVNYEPLVPSEHSCTEGGKILIESANPLENAQDQSAVRIEKGSLRFLAL 3971
QY 2288 -----LPVTLVRPLRDKIAMEKHGV----- 2308
Db 3972 EKVLLKEHSDNVMPDQIIBSKREPVAIKVQEVQGRDILLSKESLLSGIPEEORLN 4031
QY 2309 LECQVSRA-SAQV-----RW----- 2322
Db 4032 LKIQICRALQAASEQPLFSEMLNIEKVEZAVNITQEPHIMCMYLVTSAKSVTEE 4091
QY 2323 -----FKGSS 2327
Db 4092 VTIIEEDVPMANJMBELDALCAIIEEDIDILTABGPRIQOQAKTSLQEMDSFGSQ 4151
QY 2328 ELQPKYELSDGLYRKLIISDVHAEDDYTCADGV----- 2366
Db 4152 KVEPITEVESKYLISIEEVSYNVQSRVKY-LDAPVTKGVASAVVDSKQESLKPS 4210
QY 2367 ---KTSQOFFVEOSIIVIGLQDVTWBPAPA-----WFCETSIISVR 2408
Db 4211 EKESSSESGTEEVATVKIQEAGGLIKEBPMIHTPLDVTVSEBGDIHALTTSITNAK 4270
QY 2409 PPKLLGLTQAGNVGLEGETVHRLMLAR-----TCTMTGPPVHFTVGKRS 2458
Db 4271 EVNWTFENKLVPSDEKFKCLOQONTYTLVIDKNTEDHQGYVEALNDS-----GKTAT 4325
QY 2459 SARL-WVSDIPVVLTRPLEPTGRELQSVLSCDFRPAPKA-VQMYKDDTPLSPSEKPKM 2516
Db 4326 SAKLTWVRAAPVKKIEPLEVALGHLAKFTCEIQSAPNVRFQWFRAGRIEYSDKCSI 4385
QY 2517 SLEGQVABLRLMLPADAGVYRCOAGSAHSSTEVTVZAREVTVTG-----PLQDA 2567
Db 4386 RSSKYISSLEILRTQVDCGEYTCASNEYGSVCTA-TLTVTPVGGKKVRLPKPKP 4444
QY 2568 EATEGWSAFSCELSHEDVEEWSLNGMPLYNDSFHEISHKGRHRTLVLKSIQADAGIV 2627
Db 4445 EPKEB--VVLKSVLAKRPEEBEPKPKL-----EKVKPAPVPEPPPPKPVSEVPTV 4497
QY 2628 RASSLVKSTSARLEVRVPVFLKALDLSAERGTALQCEVSDPE-----AHVV 2678
Db 4498 TKRERKIPETKVP-EIKPAIPAPPEKPKPEAEVKTIKPPPEPEPTPIAAPVTPVW 4556
QY 2679 WRKDGVLGSPDKYDFLHTAG-TRGLVVDVSPEDAGLYTCHVGS-----EETRARVRV 2731
Db 4557 GKAAEK---APKEEAAPKPGIKGVPKTTPSIEARRKLRPGSGGKPPDEAPFTYQL 4613
QY 2732 HDLHVGIITKRLKTMVLE-----GESCSFECVLSHESADPAMWTVGGTKVSSSRFPQATR 2787
Db 4614 KAVPLKFVKEIDIILITSEFVGSSAIFELVSPSTAI--TTWMKDGSNIRESPKHFIA 4671
QY 2788 QCRKYLIVREAPADGAGVVSVR-----GLTSKSLIVREPAALIKPLEDO-WVAPGE 2842
Db 4672 DGKDRKLHIIDVQLSDAGYTCVLRLGNKEKTSIAKLVEELPVRFVKTILEEVTUVKGQ 4731
QY 2843 DYELRCELSRAGTPVHXLDRK-AIRKSKQYDVVCEGTWAMLVIRGASLKDAGEYTCVE 2901

4732 PLYLSCELNKE-RDVVVRKDGKIVVEKPGRIYVPGVIGLMBALTIINDADDDTAGTYVTVE 4790
2902 ASK-----STASIHVEKANKCFTEELTNLQVBEKGTAVFTCKTEHPAATVTRKGLLELRA 2957
4791 NANNLESCSCVAVVEIRDWLVKPIRDQHVKPKGTAFACDIKADTPNKNKFGYDEIPA 4850
2958 --SGKHQPSQEGTLRLTISALEKADSDTYTCDIGQAQSAOLLVQGRVHIIEDEDVD 3015
4851 EPNDKTEIIRDNHLYLKNAMPEDIAEYAVIEGKYPKAKTLIGEREVELLKPIEDVT 4910
3016 VQEGSATPRCISRPANYEPVHWFDKTPLHANEINEIDAQPGGYHVLTLRLALKDSOT 3075
4911 IYKESASPDABEISEADI-PGOWKLKGLLRSPTECEIKAE-GGKRFLTLHKVLQDQAGE 4968
3076 IYFEGDQASALRVTEKPSVFSRELTDATITEGEDTLVACETSTCDIPMCWTKDGKTL 3135
4969 VLYQALNAITALTITVKEIELDFVPLKOVTVPERQARPEC-VLTREANVWSKGPDI 5027
3136 RGSARCOLSHEGHRAOLLITGATLQDSGRYKCEAGGACSSIVRVHARVRFQEAUKDLE 3195
5028 KSSDKFDIADGKKHILVINDSQFDDDEGVYTAEBEGKTSARLFTVIRLKEMSPLEDQT 5087
3196 VLEGGAATLRCLVSSVAAPVKWCYGNVLRPGDKYSLROEGAMLELVVNRNLRPQDSGRYS 3255
5088 VKEGETATTVCELSHEKGVVWFKNDAKLHTSRVTLISSEGKTHKLEMEVTLDDLSQIK 5147
3256 CSFGDOTTGATLTVTALPAQFICKLRNKEATEGATATLRCELSKTAPVWRKGSETLRDG 3315
5148 AQVKELSSAQLKVLADPYFTVKLHDKTAVEKDEITLACEVSKOVVWFKDGEIIVPS 5207
3316 DRYCLRDQAMCBLQIRGLAMVDAABYSCVCGERTSASLTIRPMPAHFIRLHQESIE 3375
5208 PKYSIKADGLRILKIKKADKDKGEYVDCGTDKTKANVTVEARLIEVEKPLYGVEFV 5267
3376 GATATLRCELS-----KAAPVWRKGRSLRDGDRHSRLRQDQAVCELIQICGLAVA 3425
5268 GETAHEIEIUSEPDPVGGWKLKQPLTASPDCEIIEGKH-----LILHNCQLGM- 5319
3426 DAGEYSCVCGERTSATLTVPKALPAKFTGLRNEEAEGATAMWLCELSKAPV-EWRKG 3484
5320 -TGEVSFOANAKSAANLVKELPLIFITPLSDVKVFEKDEAKFECEVSREPRTFRLKG 5378
3485 PENLRDGRDRIYLRQETRCLOICGLAMADAGEYLVCVCGERTSATLTIRALPARTEDV 3544
5379 TQBITDGRDFELTKDGTGKHSWIKSAFDEAKYMFABDKHTSGKLIIEGRLKFLTPL 5438
3545 KNOEAREGATVLLQCEL-NSAAPVWRKGSSETLRDGRYSLRQDGTCKBLOIRGLANADT 3603
5439 KDVTAKEKSAVTVVELSHDNIRVWFKNDQRLHTTRSVSMODEGKHTSITFKDLSIDDT 5498
3604 GEYSVCVCGERTSAMLTVRALPIKFTGLRNEEAEGATAVLRCELSKM-APVEWVKGHE 3662
5499 SQIRVEAMGMSSEAKLTVLEGDPYFTGKLQDVTGVEKDEVILQCEISKADAPVWFKDGK 5558
3663 TLRDGRHSRLRQDAGCELOIRGLVAEDAGEYLCMCKERTSAMLTVRAMPKFIIEGLAN 3722
5559 EIKPSKNAVITDKKRMILLKALKSDIGOYTCDCGTDKTSKLDIEDREIKLVPLHS 5618
3723 EE----- 3724
5619 VEVNETARFETEISEDDIHANWKLKGEALLQTPDCEIKEEGKIHSVLHNCRLDQGTGG 5678
3725 -----ATEGDTATLWCELS-KAAPVWRKKGHETL 3752
5679 VDFQANVKSASHLRVKPRVIGLLRPLKDVTTAGTATFDCBSLVEDIPVEWYLGKGL 5738
3753 RQDRHSRLRQDGRCELQIRGLAVVDAGEYSCVCGERTSATLTIRALPARFIEDVKNOE 3812
5739 EPDSKVVRSEGVHTLTLRDVKLEDAAGEVQLTAKOFKTHANLFVKPEPPVEFKPLEDQT 5798
3813 ARGGATAVLOCELSKA-APVEWVKGSSETLRGGDRYSLRQDGTCELOIHLGLSVADTGEYS 3871
5799 VEEGATAVLECEVSRENKVKWFKNGTTEILSKKXIYVADGRVVRKLVHIDCTPEDIKTYT 5858

Db 7731 TRWKK--AGKTAGDPCDNRVTDVIEGTEVQ--FOVRAENAGVCHPSEPTEIISIEDPTS 7786
QY 5820 PPPPLQHYLE-----QVPER-----VQYQALLKELIRNKA 5851
Db 7787 PPSPPDLHVTDAGRKHIAIAWKEPKNGSGPIIGYHVEMCPVGTCKWVRNSRPKDLK 7846
QY 5852 RNRQNCALLEQAYAVVSALPORANKLHVS-----LMENYPG-----TLEALGEPTRQG 5900
Db 7847 FKVBEGVVPDKXY-----VLRVRANAIGVSPSEISENVAKODPCDKPTID-----LETH 7897
QY 5901 HFIWEG-----APGARMPWKGHNRHV-----FLFRN-----HLVICKP-RRDS 5938
Db 7898 DIIVIEGKLSIPVFRAPVPVTVSWHKGOGKVKASDRLTKMNDHISAHLEVPKSVRADA 7957
QY 5939 RTDIVSVFRNMKLSLIDNDQVEG-----DORAEV-----WQ----- 5973
Db 7958 GIYTIIT--LENKGSATASINVKVIGLPGPKOIKASDITKSSCKLTWEPPEPDGGTPIL 8015
QY 5974 ----EREDSVRKYLLQARTAIKGSW-VKEIC-----GIQORLALPV- 6010
Db 8016 HVLERREAGRTIPIVWVSGENKLSWTVKDLI PNGEYFRVKA VNKVGGGEYIELKNPVI 8075
QY 6011 ----WRPDP--FEEELADCTAELGETVKLACRVGTGPKPVLSWYKDGKAVQVDPHHILIE 6064
Db 8076 AQDEKQPPDPVVDVEVHNPTAE-----AMTITWKPLL--YDGSXIM--GYIIEK 8121
QY 6065 DPDGS-----CALILSDLTGVDSG--QWCFPAASAAGNCSTLKGILVQVPP-RF 6110
Db 8122 IAKSEERKRCNEHLVPLITLTAKGLEGEKEYQFRVRAENAG-----ISEPSRATPTTKA 8177
QY 6111 VNKVRASPV-----EGEDAQFTCTIEGAPVQIRWYKD-----GALLTT 6150
Db 8178 VDPIDAPKVVILRTSLVKKRGEIADIASISGSPYPTITWINDENVIVPEIKKGAAPLVR 8237
QY 6151 GNKFQ-----TLSEPSGLLVIVRAASKEDGLYECLELVNRLKSARASA 6195
Db 8238 RRGGEVQEBEFPVLPULTORLIDNSKGESQLRVRDSURPDHGLYMIKVENHGIAXKAPC 8297
QY 6196 ELRIGS-----PMLQAOQCHREQLVAAYE-----DTTLERADQE-----VT 6232
Db 8298 TVSVLDTGPPINFEVDIRKTSVLCKWEPPLDGGSSEIINVTLEKDKTKPDSEWIVWT 8357
QY 6233 SVLK-----RLLOPK-----ARGPS-----TGDLTGPCPCP----- 6258
Db 8358 STLRECKYSVTKLIEGKYLFRVRAENRFGPPPCVSKPLVAKOFFGPPDAPDKPVEDV 8417
QY 6259 -----RGAPAL----- 6264
Db 8418 TSNSMLVKNNEPKDNGSPILGYWLEKREVNSTHNSRVNKSLLNALKANVDGLLEGLTVYF 8477
QY 6265 ----QBTGSOPTVTGSEAPVPPRVPOPLHHEGPEQBEAIAQAQSWTVPIRMGAAPWG 6321
Db 8478 RVCAENAAGP---GKFSPPSDPKTAHDPISPGPP--PIPRVTDTSSTTIELEWEPFAPNG 8532
QY 6322 AG-----TQELLDW-----VSHSVVRETTQ-RTYTYQAIDTHTA----- 6354
Db 8533 GGEIVGYFVDKQJLVGNKNWSRTEKMKVQRTVKEIREGADYKLRVSANVAAEGGPGE 8592
QY 6355 -----RPPSMQVTIE---DVQAOGTGTAQFEAIEBGPQPSVTVYKDSVOLVDST 6401
Db 8593 TQPVTVAEQPEPAVELDVSVKGGIQAAGKTLRIPAVTVGRPVPTKVTKKEGEL-DKD 8651
QY 6402 RLSQOQEGTYSVLVRHVASKADAGVYTCLAQNTGQCVLCKAELLVLGDNBP---DSEK 6457
Db 8652 RVVIDNVGPKSLLIKDALRKDHGRVITATNSCKSGFAAARVEVF---DVPGPVLDLKP 8708
QY 6458 QSHRRK--LHSEFVEKBEIGRGVFGVKRQVQKGNKILCAAKFIPLRSTRQAAYRREDI 6515
Db 8709 VVTRKXNLLNWSDEPDGSGSITGFI--IERKDAK----- 8743
QY 6516 LAALSHPLVTGLLDQETPKTKLILILELCSSEBLDLRLYRKGVVTEAEVKVIOQLVEGL 6575

Db 8744 -----HTWROPIETERS-KCDITGLLEBQ 8766
QY 6576 HYLHSHGVL-----HLDIKPSNILLMVHAPAREDIKICDFGPAQNTIPAEQLQFSQYGS 6626
Db 8767 ET--KFRVIANKFKGCGPPEIGP--ILAVDP-----LGP 8797
QY 6627 PFEVSPETIIQQNPVSEASDIWAMGVISVLSLTCSSPFAGESDRATL-LNVLEGRVSWSSP 6685
Db 8798 P--TSPF-----RLTYT-----ERQSTITLTDWKEPRNGGSP 8828
QY 6686 MAALHSE-----DAKDFIKATLQAPQAPSAQAQCLSHPWFLKSMFAEBAHFINTCOLKFL 6741
Db 8829 IOGYIIEKRHKDPDF-----ERVNKLCTPTTFLVNLDEHQMRYFRVK----- 8873
QY 6742 LAERSMQRSLMSYKSLVNMRSIPELIRGPPDPSPFSGVARHLCDRTGGSSSSSSSDNEIA 6801
Db 8874 -----AVNEIGESEPSLP-LNVV-----IQDDEVP 8897
QY 6802 PFARAK-----SLP--PSPVTHSPJ-----LHPRGFLRPSASL---PEBAE 8837
Db 8898 PTIKRLSVRGDTIKVKAGEPVHI PADVTGLMPKIEWSKNETVIEKPTDALQITKEEVS 8957
QY 6838 ASERSTEAPAPPASPEGAGPPAAQGCVRHRSVIRSLFYHQAGESPETHGALAPGSRHPAR 6897
Db 8958 RSEAKTELIPKAVREDKGTIYVTASNRLGVSFRNVHVEYDR-----ESPENLAV 9009
QY 6898 RHLLKGGYIAGALP--GLREPLMEHRVLEBAAREQA-----TLAK-----A 6940
Db 9010 TDIKAESCYLTDAPLDNGGSE--ITHYVIDRDASRKKAEEVEVTNAVEKRYIWKLI 9067
QY 6941 PSPETALRLPASGTHLAPGSHSHLEHDSPTPRPSEACGEAQRPLSPAGSGAPIRDMGH 7000
Db 9068 PNGQYFRVAVNKY---GISDECKSDKVIQDP-----YRLGPP--GKP-KVLAR 9113
QY 7001 PQGSKQLPST-----GGHFGTAQ--PERPSPDSF-MQOPAPFFCHPKQG-----SAPQ--E 7045
Db 9114 TKGSMLVSTPPLDNGGSPITGVWLEKREEGSPYMSRVSRAPITKVGLKGVFNVRILLE 9173
QY 7046 GCS-PPHVAVAPCPGFPFGSGCKEAPLVFSSP-FLQOAPAPAPAKASPLDSEKMGPGDI 7103
Db 9174 GVKYQFRAMAINAAGTGPPE-----PSDEVAGDPIFFPGP-PSCDEVKDKT-KSSI 9224
QY 7104 SLPCRPKPGPCSPGSAQSSQVSSLRVGSQVGTBEPGPSIDARGM--TOEAEPLSDS 7161
Db 9225 SLGKWP--PAKGG-----SPIKGIYVMEQECT-----TDKRVNPEPKLIIT 9266
QY 7162 ----TPTLQRPQOATWRKFSLGGRGYAGYGTFAFGAGMGLGQGPWAPARAWAV 7217
Db 9267 CECVWPNLKE-----LRKYRF-----RVK-AV 9287
QY 7218 SQSEEEQOEAREASEOQOEAREABSPLOVSARVPVEVG-----RAPTRS 7264
Db 9288 NEAGESEPSDTTGEIPATDIOEB-----PEVFI---DIGAQDCLVCKAGSQIRIPAVI 9337
QY 7265 SPBFTP---WEDIGQVSLVQ---IRDLSGDAEAADTISLDSIVDPAYNLNLSLYDI--K 7316
Db 9338 KGRPTPKSWEFPDGKAKAMKOGVHDIPEDAQLETAENSSVIIIPBCKRSHTKYISITAK 9397
QY 7317 YLPFSFVI FRKVPKSAQPEPPSPMAEELAEFF-EPTWMPGELG----- 7360
Db 9398 NKAGQKTANCRVKMVDVPGPKDKLVSDITRSCRLSWKMPDDDDGDRKGYVIEKRTID 9457
QY 7361 -----PHAGL-----EITBESEDVDALLAAEAAGVRKRKWSPPSRSLFHPGRHLPL 7406
Db 9458 GKAWTKVNPDCGTTTFVVPDLLSEQQYFRVRAENRFGIGPPVETIQTARTDP--IYPP 9515
QY 7407 DEFAELGLRERVKASV-----PHISRLKGRPEGJEKEG----- 7440
Db 9516 DPTIKLIGITIKNTVHLSWKPKNDGSPVTHYVECLAWDPTGTKEAWRQCNKRDE 9575
QY 7441 -----PPRKKPGLASFRLSGLKSW-----DRAPTF-LRELSDETIVL 7476
Db 9576 ELQPTVEDLVEGGEYFRVKA VNAAGVSKPSATVGCDCQPDMPFSDIKFERWE--VEE 9633

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: September 13, 2004, 11:13:38 / Search time 148 Seconds
(without alignments)
5178.744 Million cell updates/sec

Title: US-10-077-130-5
Perfect score: 41273
Sequence: 1 MDQPSGAPFLTRKAFV.....RNREKRALLYKHNLAQVR 7968

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 3577.5 | 8.7 | 26926 | 1 | I38344 |
| 2 | 3313 | 8.0 | 7962 | 1 | I38346 |
| 3 | 2629 | 6.4 | 4162 | 2 | T42633 |
| 4 | 2489 | 6.0 | 6642 | 2 | T29757 |
| 5 | 2048 | 5.0 | 5198 | 2 | T32290 |
| 6 | 2043.5 | 5.0 | 5175 | 2 | T20992 |
| 7 | 1636.5 | 4.0 | 7160 | 2 | T27935 |
| 8 | 1634 | 4.0 | 6839 | 2 | S57242 |
| 9 | 1619 | 3.9 | 6831 | 2 | A88852 |
| 10 | 1392 | 3.4 | 6658 | 2 | T13931 |
| 11 | 1074 | 2.6 | 4391 | 2 | A38096 |
| 12 | 1045.5 | 2.5 | 6805 | 2 | S20901 |
| 13 | 1012 | 2.5 | 1906 | 1 | S68235 |
| 14 | 963.5 | 2.3 | 2783 | 2 | T34416 |
| 15 | 869 | 2.1 | 1323 | 2 | PN0568 |
| 16 | 786.5 | 1.9 | 3488 | 2 | T34418 |
| 17 | 767 | 1.9 | 3707 | 2 | S18252 |
| 18 | 703.5 | 1.7 | 1398 | 2 | T25568 |
| 19 | 645.5 | 1.6 | 1176 | 2 | JN0583 |
| 20 | 622 | 1.5 | 1147 | 2 | A59307 |
| 21 | 571 | 1.4 | 1274 | 2 | S55050 |
| 22 | 532 | 1.3 | 1694 | 2 | S50065 |
| 23 | 528.5 | 1.3 | 1132 | 2 | A35089 |
| 24 | 520 | 1.3 | 1142 | 2 | T36845 |
| 25 | 519 | 1.3 | 3375 | 2 | T19821 |
| 26 | 501.5 | 1.2 | 608 | 2 | A35021 |
| 27 | 480.5 | 1.2 | 610 | 2 | T28798 |
| 28 | 472 | 1.1 | 1896 | 2 | T08851 |
| 29 | 465.5 | 1.1 | 2541 | 2 | T29340 |

RESULT 1

I38344

titin, cardiac muscle [validated] - human

N;Alternate names: Connectin

N;Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)

C;Species: Homo sapiens (man)

C;Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 15-Sep-2000

C;Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393

R;Label: S.; Kolmerer, B.

Science 270, 293-296, 1995

A;Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.

A;Reference number: A57430; MUID:96026330; PMID:7569978

A;Accession: I38344

A;Status: nucleic acid sequence not shown; translation not shown; translated from GS/EM

A;Molecule type: mRNA

A;Residues: 1-26926 <LAB1>

A;Cross-references: EMBL:X90568; NID:G1017424; PID:G1017425

R;Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.

Biochemistry 34, 553-561, 1995

A;Title: Dissecting titin into its structural motifs: Identification of an alpha-helix

A;Reference number: I38345; MUID:95119041; PMID:7819249

A;Accession: I38345

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1977-2014 <MUG>

A;Cross-references: EMBL:X83270; NID:G602579; PIDN:CAA58243.1; PID:G602580

A;Note: conformation and properties are reported for a synthetic peptide corresponding

R;Label: S.; Gauter, N.; Lakey, A.; Trinick, J.

EMBO J. 11, 1711-1716, 1992

A;Title: Towards a molecular understanding of titin.

A;Reference number: S20897; MUID:92258380; PMID:1582406

A;Accession: S20898

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 13597-14200, 'I', 14202-14696 <LAB2>

A;Cross-references: EMBL:X64698; NID:G37192; PIDN:CAA45939.1; PID:G37193

A;Accession: S20897

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 16330-16392, 'S', 16384-16756, 'F', 16758-16860 <LAB3>

A;Cross-references: EMBL:X64699; NID:G37190; PIDN:CAA45940.1; PID:G37191

A;Accession: S20899

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-

A;Cross-references: EMBL:X64697; NID:G37190; PIDN:CAA45938.1; PID:G37195

R;Kolmerer, B.; Olivier, N.; Witt, C.C.; Herrmann, B.G.; Label, S.

J. Mol. Biol. 256, 556-563, 1996

A;Title: Genomic organization of M line titin and its tissue-specific expression in two

A;Reference number: S63665; MUID:96177761; PMID:8604138

A;Accession: S63665

A;Status: nucleic acid sequence not shown

| | | | |
|------|----|---|------|
| 1505 | Db | EHQWKPMFVEXKLKNVNIKEGSRLEBMKVRATGNPNPDVLWLNQSDIIVPHKYKPKIRIEGTX | 1564 |
| 329 | Qy | -----R 329 | |
| 1565 | Db | GEAALKIDSTVSQDSAWYTATAINKAGRDTRCKVNVVEFAPEPERKLIIPRGTVRAK | 1624 |
| 330 | Qy | BPAPV-----FKRLQDLVREKESATFLICE---VP 357 | |
| 1625 | Db | EIAAPELEPLHLRYGOFQWEEGLYDKKQKOPFFKKXLTSLRKLRFPAHFECRUTPIS | 1684 |
| 358 | Qy | QPSTEAAWFKEBTRL-----WASAKYGI-----EEBGTERRLT 390 | |
| 1685 | Db | DPTWVWENLHGXPLEAANFLRMINFEFGYCSLDYGVAYSRDSGIIICRATNKYGTHTSA | 1744 |
| 391 | Qy | VRNVSDADDAYICETPEBSR-----TVAEALVQGNL-----LRKLPRKT 430 | |
| 1745 | Db | TLIVKDEKSILVEESQLPEGRGLQRIEELERMAHEGALTGVTTDQKQKPDVLVYPEV | 1804 |
| 431 | Qy | AVRVGDTAMP-CVELAVPVCPSVHMLNQBESVAGGVAISABGTRHTLTISQCLBEDVGQ | 489 |
| 1805 | Db | RVLEGTARPRCVTGYPQPKVNWYLNQILIRKSKFRVRYDGI-HYLDIVCKCSYDTGE | 1863 |
| 490 | Qy | VAFMAG-----DCQTSRFCVSAPRKPPLQ-----PVD 518 | |
| 1864 | Db | VKYTAENPEGVIEBHKVKEIQOQREDFRSLRRAPEPRPEFHVHEPGKLOFEVQKVDPRVD | 1923 |
| 519 | Qy | -----PVVKARMESSVILSWSPPHGE-----RPVTDGY-----LVEKKKLTJTYW 560 | |
| 1924 | Db | TTTEKVEVVLKRAERITHEKVPSESELBSKFKRRTEEGYEAITAVELEKSKKDSYEBE | 1983 |
| 561 | Qy | I-----RCHAEAWATPELTIVAD---VAERGNFOFRVSALNSFGQSPYLEFPFGTVHAPKL | 613 |
| 1984 | Db | LKRKTKDELLHW--TXELTEEBEKKALAEGBKIIPFKDKIELSPSMB-----APKI 2034 | |
| 614 | Qy | AVRTPKAVQAVEGGEVTSVDLTVASAG-----EWFLDGQALKASS-VYEHCDORTRH | 666 |
| 2035 | Db | FERIQSQTVG--QGSDAHFRVR-----VGKPDPECEWYKNGVKIERSDRIYWYEDNVC | 2088 |
| 667 | Qy | TLTIREVPASLHGAOLKF---VANGIESSIRMEVRAAPGLTANKPAPAAAAREVLAARLHE | 722 |
| 2089 | Db | ELVIRDTAB-DSASIVKAINIAGETSSHAFLLVQAKLIITFQ-----ELQDVAKED 2143 | |
| 723 | Qy | E-AOLLAEUSDQAAAVTWLKGRTLSFGPKYEVQASAGRRVLLVRDVARDDAGLYECV-- | 779 |
| 2144 | Db | TMATFECETSEPPVKVWKYDGMGEVHEGDKYRMHSDRKVHFSLILATIDTSDAEDYSCLV | 2203 |
| 780 | Qy | -SRGGRIANQLSVOGLARFLHKDMAGSCVDVAGGPAQPECETSEAHVHVHHYKDGWELG | 838 |
| 2204 | Db | EDENVKTTAKLIVEGAAVFVKELQD--LEVPSYSGELEICIVSPNIEGKVHNDVELK | 2261 |
| 839 | Qy | HSGERFLOQDVGPRHRLVAATVRQDEGYSCKRVGSDSDVDFRLRVSEPKVVFVFAKEQLARR | 898 |
| 2262 | Db | SNKYITITSPRG-RQNLTKVDTKEDQGEYSFVIDGKTCTKLM-KPRPIALQGLSDQ | 2319 |
| 899 | Qy | KLOAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKYCMNEATGCTTRLVVOQAQADAG | 958 |
| 2320 | Db | KVC--EGDIVQLEVKVLSLEPSVEGVMMKQGEQVQPSDRVHIVIDKQGHMLLIEMTKEADG | 2377 |
| 959 | Qy | EYS-----CEAGGORLGFHLDDKPEKPVVFAKDQVAHSEVQAEAGANATLSCEVAQAQ- | 1010 |
| 2378 | Db | NYSFTTIPALGLSTSGRVSYVSDVITPL-----KDVNVIETKAVLECKVSPDV | 2427 |
| 1011 | Qy | AEVWYKDGKLSSSLKVVHAEKGBRRRLVVOAQKTDAGDYSCBARGQVRSFPLHITEP | 1070 |
| 2428 | Db | TSVKYTLNDEQIKPDRVQAIKVGTKQRIVINRTHASDEGPYKLIIVGRVETNCLNSVEKI | 2487 |
| 1071 | Qy | KMMFAEQSVHNEVQAEAGASAMLSCVEAQAQTEVTWYKDGKLSSSKVGMEVKGCTRR | 1130 |
| 2488 | Db | KIIRGLREDTCTETQ-----NVVFEVLESHSIDVLWNPDKKEIKPSSKYKIEAHGKIYK | 2542 |
| 1131 | Qy | LVLPOQAKDAGEYSCBAGQQRVSFHLHITEPKGVFAKBCQSVHNEVQAEAGTAMUSCEV | 1190 |
| 2543 | Db | LTVLNMMKDEGKYTYFAGENMTSGKL--TVAGGAIK--PLTDDTVASS-QBAVPECEV | 2597 |

Db 4263 TTGDTCTLECTVA-GTPBELSTKWFKDGKEL-TSDNKYKISFFNKVSGLKLIINVAPSDSGV 4320
Qy 4504 VTFACRDV---ASALTVLGLPDPEDA-----EVAHSHHTVTLGW 4543
Db 4321 YSFVQNPVGKSDCTASLQVSDRTVPSPFTRKLKETNGLSGSSVWMECKYVGPPIVSVM 4380
Qy 4544 AAPMSDGGGLCGVRVVEKGATGQWRLCHELVPGPEVVVDGLAPGETYKFRVAAGVPVG 4603
Db 4381 ---FHGN-----BISGRKYQTLTNTWC---ALTVMLEBDSGDTYCIATNMAG 4426
Qy 4604 AGEVHLFQTVRLAEPKPPVPPSAPESQVAAAGEDVSLELEVAEAG-EVIWHKMER 4662
Db 4427 SDE-CSAFLTVR-EPPSFV---QKDDPMDVLTGNTVFTS-VKGTPPSPSVWFKGSS 4479
Qy 4663 IQPGREVVSQGOQMLVKGFTAEDQGEYHCGLAGSGICPAAAFQVALSPASVDEAP 4722
Db 4480 LVPDRCNVLEDSVAELELPDVTDSQSGYTCIVS-----NEAG 4519
Qy 4723 QPSLPPAAQEGDLHLHLEALAR---KRRMSREPTLDSISELPEEDGRSQRLOPQEAEEVAP 4780
Db 4520 KASC-----TTHLYIKAPAEVKE-----LNDYSIEKKG-----P 4549
Qy 4781 DLSEGYTADELARTGADASHTSSDDESAGTSPSLVYLYKAGRPOTSPASKVGPAA 4840
Db 4550 LILEGTFT-----GTPPISVTVKXNG----- 4570
Qy 4841 PSVKPQQQBEPLAAVRPLGLDSTKDLGDPMSMDKAAVKIOAFKGYKVRKEMKQEQPMF 4900
Db 4571 INVTPSORCNITTEKSPILPSPSTVEDAGQNCYIE-NASGKD-SCSAQIILIPPPYF 4628
Qy 4901 SHTEGDEAQVGDALRLCEVVASKADVRARMLKDGVEL---TGRHHHIDQLDGTGCSLL 4957
Db 4629 VKOLEPVKVSVDGSASLQCLAGTPEIGVSWYKGDTKLRPTTYTKM-----FRNNVATLV 4684
Qy 4958 IAGLDRADAGYTCQVSKKQVTHSACVVVSGSEAESSGSELDDAFERRAARLHRL 5017
Db 4685 FNQVDINDSGEYICAKNSVGEVSASTFLTVQEO-----KLPSPSRQLRDVQE- 4733
Qy 5018 PRTKSPAEVSEDEHFLGADEGPAPPEFPADQTYREDEHFICIRFEALTEARQAVTRFOE 5077
Db 4734 -----TVGLPVVFDCAISG-----SEPIVSWYKDG-----KPLKDSNPVQTSFLD 4774
Qy 5078 MFATLGI-GVEIKLVEQPRVEMCISKEPAPVVPPEPLPSLLTSDAAVFLTELQOE 5136
Db 4775 NTATLNIKTRSLAQO-----YSCA-----TNPIGS--ASSARLILTEGNKP 4818
Qy 5137 VQD-----GVPVSFDCVVTGQPMPSVRWFKDGKLEEDDHMYNEDQOQGHQI 5186
Db 4819 FFDIARLAPVDVAVGESADFECHVTGTQPIKVSNAKDSREIRSGKYQISYLENSAH-LTV 4877
Qy 5187 TAVPADMGVYRCLAENSMVSGSTKAELRLVDLTSTDYDTAADATESSSYFSAQYLSRE 5246
Db 4878 LKVDKDGSGQVTCYAVNEVGKDSCTAQLNI-----KE 4909
Qy 5247 QEGTESITDEGOLPOVVEELRLQVAPGTRLAKQL--KVKGYPAPRLYMFKQGOPLTAS 5304
Db 4910 RLIPPSFTK--RLSTVEE-----TEGNSFKLGRVAGSQPIVAVYKNNIEIQPT 4958
Qy 5305 AHIRMTGKILHLEIISVTRSDSQYAAIYISNANGAYSSARLLVRGPDEPEKPSADV 5364
Db 4959 SNCEITFNKNTLVQVRKAGMNDAGLYTKVSNDAAGSALCTSSIVIK---EPKPPPVFDQ 5015
Qy 5365 HEQLVPPRMLEFTRPKYKKSITFSVKVEGRPVPTVHMLREAEARGVLMIGP-DTPGY 5423
Db 5016 H-----LTFVTVSEGEYVQLSCHVQSGSEPIRQWLKAGRE-----IKPSDRCSF 5059
Qy 5424 TVASSAQOHSVLVLDVGRHQGTTCIASNAAGQALCSALHVSGLPKVPEQEKVEALI 5483
Db 5060 SPASGTA--VLELRDVAKADSGDYCKASNVAGSDTTKSKVTIKDKPAVAPATKAAVDG 5117
Qy 5484 STFLOGTQAISAQLEITASF-ADLGQORKEPELA-AKEALGHLSAEVGTBEFLQKLTS 5541
Db 5118 RLFFVSEPOSIRVWEKTTATFTAKVGG-----DPIENVKWTGKWRQLNQGRVFIHQ----- 5170

Qy 5542 QITEMVSAKITQAKLOVPGGSDSDSKTPS-----ASPRHGRSRPSSIO-----ESSSES 5592
Db 5171 -----KGDEAKLEI-----RDTIKTDSGLYRCVAFNEHGEIESNVNLQVDERKKQEK 5217
Qy 5593 EDGDARGEIPIFYVVTADYLPFGAEQDAITLREQOYVEVLDAAPHLRMLVTRTKPTKSSPS 5652
Db 5218 IEGDLRA-----MLAKTPIKKGAGEE-----EEDIMELLKNVDP----- 5253
Qy 5653 RCGWSPAYLDRRLKLSPEWGAAPPEPPGAEVSEDEYKARLSSVIOELLSSEQAF--VE 5710
Db 5254 -----KEYEKYARMYG-ITDFRGLLOQFELK 5279
Qy 5711 ELQFLQSHHLOHLERCPHVPIAVAGQKAVIFRNVRDIGRPHSSFLQELQOCDTDDDVAMC 5770
Db 5280 OSQBEETHRLE-----IEETERGERD----- 5300
Qy 5771 FTKNOAAFEQYLEFLVGRVQAEVVVTAIOEFYKYAEAEALLAGDPSQPPPLQHLYLE 5830
Db 5301 -----EKEFEELVSFQQEL-----SOTEP----- 5320
Qy 5831 QPVERVQVQALLKELIRKARNRONCALLQEAQAVVVSALPQAEAKLHVSIMENYPGTL 5890
Db 5321 -----VTLIKD-IENQTVLKNDVAFE-----IDIKINYP-- 5349
Qy 5891 EALGEPIRQHPIVWEGAPGAMPWKGNRHVFLFRNHLVICKPRRDSRTDTVSVFRNM 5950
Db 5350 ----- 5350
Qy 5951 MKLSSIDLNDQVEGDDRAFEVWQEREDSVRKYLLQARTAIKSSW-VKEICQIQORLALP 6009
Db 5351 IKLSWYKTEKLEPSDK-PEISIDGD-----RHLRVKNCOLKQDQNYRLVCGPHIASAKL 5405
Qy 6010 VWRPDPFEBELADCTAELGETVKLACRVT---GTPKPVISWYKDGAKVQVDPHHILIEDP 6066
Db 5406 TVIEPAWERHLQDVLKEGQT---CTMTVQSPVNVKSEWFRNGRILKPOGRH-KTEVE 5460
Qy 6067 DSCALILDSLTVGSDQYMCFAASAGNCSTGLKILVQVPP-RFVNVKRAFPFVEGEDA 6125
Db 5461 HKVHKLTIAVRAEDQGOYTC---KYBLETSAELRIEAEPIQTKRQNIQNVSEHQSA 5516
Qy 6126 QFTCTTEGAPYQIIRWYKDGALLTGNKFQTLSEPRSGLLVLVIRAAKSEDLGLYCEELV 6185
Db 5517 TECEV-SFDDAIVTWYKGTPELTESQYKYNFNDGRCH--YMTIHNVTPDDEGVY--SVI 5571
Qy 6186 NRL---GSARASAEIRIQSPMLQAOBQCHREQLVAAVEDTTLERADQEVTSVLKLLGPK 6242
Db 5572 ARLEPRGEARSTAEIYL-----TTKEIKLELKP- 5599
Qy 6243 APGPSGDLTGPGP-CPRGAPALQETGSOPTVTGTSEAPVPPVPPQPLLHEGPEQPEBA 6301
Db 5600 ---PDIPDSRVPIPTMPTRAVPPEL---PPVV---APPVPLLPTP-----EEKPP- 5643
Qy 6302 IARAQEWTVPIRMEGAAPFAGCTGELLMDVHSHVRETTQRTYTYQOAITHTARPPSMQV 6361
Db 5644 -----PKRIE-----VTKKAVKQAKV-----VAKPREM-- 5668
Qy 6362 TIEDVQAGTGTAQFAIIEGDPQSPVWTKDSVQLVDSTRLSQOEGTYSVLVR--HV 6419
Db 5669 -----TPREIIVKXPPPTLLIPAKAPEIID---VSSKAEVFKMTITRKKEV 5713
Qy 6420 ASKDAGVYTCLAQNTGGVLCCKAEALVGLGDNEP--DSEKQSHRRKLHS-FYEVKEEIGR 6476
Db 5714 QKEKEAVY-----EKKQAVHKEKRVFIESFEPEYDELEVEPTEPEQPYVEEDDEYE 5767
Qy 6477 GVFGFVQRVQHKNKILCAAKFIPLRSRTRAQAYBERDILAALSHPLVTGLLDQFTRKT 6536
Db 5768 EIKVEAKKEVHEWE-----EDFEEGQFYERE-----EGYD----- 5799
Qy 6537 LILILECSSEULLDLRYKGVVTEAEVKYIQQOLVEGLHYLHSH-----GVLHLDIKP 6590
Db 5800 -----EGEEWEAYQEREVIQVQKEVYES-----SHERKVPKVPKPKAPP 5841


```

QY 6591 SNILMVHPAREDIKICDFGPAQNTIPAELOQSQVSEFVSPELIQNPVSEASDIWAMG 6650
Db 5842 PKVIKKKVIKIE-----KTSRRBEEKVQVTKV--PE-VSKTIVPOKP-----5883
QY 6651 VISYLSLTCSPFFAGESDRATLLNVLEGRVSWSPMAAHLSADAKDIKATLORAPQARP 6710
Db 5884 -----SRPVPQEE-----VIEVKVPAVHTKXWVISEKQKPFASHTSEVSVTP 5927
QY 6711 SAAQCLSHPWFLKMPABEAHPINTKQLFLLARSQRSLMSYKSLVMSRISPELRLGP 6770
Db 5928 EV-----QKEIVTBEKIHV--AVSKRVEPP 5950
QY 6771 PDSPLGVARHLCDTGGSSSSSSSSSDELAPFAKSL-PPSPVT-----6815
Db 5951 PKVPFL-----PEKPAEEVAVPPIPKVFPAPKVEVPKVPBEKPK 5995
QY 6816 -----HGPLLHPRGLRPSASLPEEASERSTEAPAPASPEGAGPPAAQGCVPRHSVI 6870
Db 5996 VPVPKGPAPAPKVPKVPBEKIPVPVAKKEAPPAKV-----PEVQKGVVTEKI- 6050
QY 6871 RSLFVHQAQSEHGA-----LAPGSRHPARERHLLKGYTAGALPGRLPELMHRV- 6923
Db 6051 -----TIVTQRESPPAVPEIPKVKVPBEKVPKVEEVPPPPKVPALP--KKVPVBEKVA 6106
QY 6924 -----LEEAAREQATLLAKAPSFETALRLPASGTHLAPGH--SHSLEH 6966
Db 6107 VPVPVAKKAPPPRAEVSKTIVVEKRFVAEKLKSFVAPQVRETVRHEVSAEESVSEER 6166
QY 6967 DSPS-----TPRSSACGE-----KQRLPSAPSGGAPIDMG 6999
Db 6167 EGVSVSVREBEREEBAEVTEYVMEPEEYVVEEKLHIISKRVAEPA-----EVT 6220
QY 7000 HPQGSK-----QLPSTGCHPQTAQ--PERPSPDPSWGO--PAPFCHPKQGSAPQBGCSHPHA 7052
Db 6221 ERQEKIVLKEPKI-PAKIEBPPPAKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6278
QY 7053 VAPCPGSPFPFGCK--EAPL-----VPSPPFLOQOPAPPAPAKASPPLDKMG- 7099
Db 6279 --KPVEKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6337
QY 7100 PGDISLPKPKPGPCSPGSSASQSSQVSSSLVGGSSQVGTPEGPSLDAEGWTQE--AED 7157
Db 6338 PEKVVLKLPKKEE--EPKAVTTEFRKRVKVEKUSTEAPKREPOPIKEVTIMEKERAVT 6396
QY 7158 LSDSTPLQPOQOATNKRKSLGRRGYAGVAGYGTAFQGDAGMLGQGMWARIAMAV 7217
Db 6397 LEEAVSVQREEEVEEYDYKEFEYEYPTEDYQV-----6433
QY 7218 SQSEEEQEARAESQSEEQEARAESPLQVSARVPVEVGRAP-----TRSSPE---PT 7269
Db 6434 --BEYEREVEVEEHEEYITEEKIPKVPVEEVPVTKPAPPAKVLKVAPEKVPV 6491
QY 7270 PWEDIGQVSLVQIRDSGDAEAAADTISLDISEVDPAYLNSLDYIKLFFEFMIPRKVP 7329
Db 6492 P-----KGRPEGLEKGGPPKPK 7446
QY 7330 KSAQPEPPSPWAEELAEPEPTWMPGELGPHAGL-----EITSEEDV-----DALLA 7379
Db 6495 KKLKPPPK-VPEPKVPKVEKI-----HISITKREKQVTPAAKVPKPKRVVA 6544
QY 7380 EAAVGRKRKWSRSLFHPGR--HLPLD--EPAELGLRERVKASVEHISRL-----7429
Db 6545 EEKVPVPERKAP-----PVAVPEVPKELEPEEVAFEVEV---VTHVEEVLVEEERY 6594
QY 7430 -----KGRPEGLEKGGPPKPK 7446
Db 6595 IHEEBFIITEEVVVPVVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6644

```

RESULT 3

T42633

connectin/titin - chicken (fragment)

C:Species: Gallus gallus (chicken)

```

C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C/Accession: T42633
R/Yajima, H.; Ohtsuka, H.; Kawamura, Y.; Kume, H.; Murayama, T.; Abe, H.; Kimura, S.; Ma
Biochem. Biophys. Res. Commun. 223, 160-164, 1996
A/Title: A 1.1 kb 5'-terminal cDNA sequence of chicken breast muscle connectin/titin re
A/Reference number: Z22221; MUID:96254045; PMID:8660363
A/Accession: T42633
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-4162 <YAJ>
A/Cross-references: EMBL:D83390; NID:gi513029; PIDN:BAAL1908.1; PID:gi513030
A/Experimental source: breast muscle
C/Keywords: skeletal muscle

```

```

Query Match      6.4%; Score 2629; DB 2; Length 4162;
Best Local Similarity 23.4%; Pred. No. 6e-70;
Matches 1110; Conservative 694; Mismatches 1874; Indels 1074; Gaps 172;

```

```

QY 9  APELTPKAFVSV-GKDATLSCOIVGNPTPOYSWEKDQPVTA---GARFLAQDGD 63
Db 5  APTF-TQPLQSVVALEGAATFEAHISGFPVPEVSWYRDGQVLSAATLPGVQISFS-DGR 62
QY 64  LYRLTLDLALDGSQYVCRARNAIGEAFVAVGLQVDAEAAACABQAPHLLRPTSIKRVRE 123
Db 63  A-KLVIPSVTEANSGRYTIQATNGSGQATSTAEILLVTAGTA---PPNFSQRLQSMARQ 117
QY 124 GSEATPCRCVCGSPRPVAVWSKDGRRLGEPPGPRVVEELGEASALRIARPRDGGTYE 183
Db 118 GSQVRLDVRVTGIPFPVVKFYRDGVEI--QSSPDFQILQEGDLYSLIAEAPEYDSGTYS 175
QY 184 VRAENPLGAASAAAALVVDSDAADTASRGPTSTAAALLHLQRREA-----229
Db 176 VNATNVGRATSTAEILLQGEAEVPAK-TKTVSTAQISQTEQARIEKKIEHFDARS 234
QY 230 -----MRAEGAPA-----SPPSTGTRTCTVTEGKHARLSCYVTGEPKPKETVWKQGLV 278
Db 235 LTSVEMVIEGAAQQLPHKAPPRPPR-----TSKSTPPVITAK-AQMA 279
QY 279 TEGRRHVVEDAQENFVLKILFCQSD--RGLYCTASNL-----VGQYSSVLVVRE 330
Db 280 RQQSPSPV-----RQSPSPVHVAPTPSPVRSVSPAGRISTSPIRVKS 324
QY 331 PAVPFPKR---LQDLEV-----REKESATFLCEVQPQ--STEAAWPKETRLWASAKYGI 380
Db 325 PS-PIKAAQVVTGPAEVLPPWREGYSATAEAAQMKETRVSTSAITEER-W-EGRYGL 381
QY 381 EEEGTERRLLVRNVSADDDAVYICETPEGSRTVAELAVQGNLLRKLPRKTVAVGDTAMF 440
Db 382 QEQ-----VTISGAAA-----GEVAAKAKEVRKEPEKTPV---PTVII 416
QY 441 CVELAVPGPVVHVLNRQNEVAVGRVAISAEGTRHTLTISQCCLEDYGVAFMAGDCQTS 500
Db 417 ATDKA-----KEQERISTAREBISA--RH-----EQV-----HVS 444
QY 501 TRFCVSAPRKPPLOPPYDVPVVKARMESSVILSNSSPPHGERPVTIDGILYVEKKLGTITW 560
Db 445 HEQTEAGKRAEAVATVVAADQARVS-----PWETE-----QVDETYYKKKTL-EYGY 492
QY 561 ----IRCHEAE---WVATPELTADVVAEENGFQFRVSALNSFGQSPYLEPFGTVHLPKL 613
Db 493 KEHAVKDEHAQAEHHVATKEVK-----IVVYYPE- 521
QY 614 AVRTPLKAVQAVEGEVTFVVDLTVASAGEWFLDGGQALKASSVVEIHCDRTRHTLTIREV 673
Db 522 -----KHIPAAEKKEVHVSTEIK-----RET 542
QY 674 PASLRGAQLKFVANGIESSIRME-----VRAAPGLTANK-----PPAAAAREVLARLHEAQL 726
Db 543 EAK-----IEKTIHIEHPRPTASPHFTVSKTAVPKPDHTYEVSIAGSAMATL 590
QY 727 LAELSDDAAAATVWLKQRTLSGPQKVEVOASAGRRVLLVRDVARDDAGLYCEVCSRGRIA 786
Db 591 EKLSATSAQAQKITKPKPPQLXP-HEVKIKP-----621

```


Db 2526 KKPXDVTALNANVVSFELSVSHDTPVVRWPHKVNVELKQSKYKMSQKVKHKLMLNISP 2585
 QY 2711 EDAGLYTCHVGSEETRAARVRVHDHVGITKRLTMEVLBEGSCFECVLHSHESASDPMW 2770
 Db 2586 ADAGEYAFVQLECKAKLFTETIH--ITKTKMSIEIPETKTASFQCEVSHFN--PSVM 2641
 QY 2771 TVGGKTVSSRFOATROGRKVIYLVREAPSDAGEVVSFVGLTSKASLIVREPAII 2830
 Db 2642 LKNGVEIEMSKPFIYVQKHLQJNIMNTSEDSAEYTFVCGNDRVSATLV--KPLIT 2699
 QY 2831 KPLEDQWAPGEDVELRCELSRAGTPVHWLKORKAIRKOKYDVVCGTWMAMLVIRGASL 2890
 Db 2700 SMLEDINAEKXDTITFEVTNYEGISYKWLKNGVEIKSTDKCQIRTKKLTSLIRNVHF 2759
 QY 2891 KDAGEYTCVEASKSTASLHVVEKANGFTEELTNLOVEEKGTAFTCKTEHPAATVTRK 2950
 Db 2760 GDAEYTFVAGKAASSATLYVEARHIEFRHKIDIKVEKRAIFCEIESEPVDQVQMMK 2819
 QY 2951 GLEELRASGRKHPQSEGLTLELT:SALEKADSDTYTTCDIGQASRAQLLVQGRVHIIE 3010
 Db 2820 DQELQIGDRMKIQREKYVHRLIIPSTKMSDAGQYTVVAGNTSSANLIVEGRDVRISI 2879
 QY 3011 LEDVDVQEGSSATFRCRISPANYEPVHFLDKTFLHANELNEIDAOPGY-----HVL 3064
 Db 2880 RKEIQVIERQAEIEFEVNEDDIEP-QWYKD-----GIEINFHYERYSYVVERRHRMS 2933
 QY 3065 LRQIALKDSGTIYFEAGDQASAAALRV--EKPSVFSRELTDTATTEGEDTLVLCETSTC 3122
 Db 2934 IFETTSYDAGYTFVAGNRSSVLYVNAPEPQII-QELOPTVSGKPARCAIISGK 2992
 QY 3123 DIP-MCWTKOQKTLRGARCOLSHEHRAQLLITGATLQDSGRYKCPA-----GGACSSSI 3177
 Db 2993 POPKVSVMYKDDQLSPGCKFLHDAQEYTLTLLIETFPEDSAVYVCEAKNDYGVATTSAS 3052
 QY 3178 VRHARPVRFQ-----ALKDLEVLGGATLRCVLSSVAAPVKWCYGNVLR 3225
 Db 3053 LSVIPEVVSPELPPVPPVPAVIVPLRDVATSEQASRFQCVGTGTLKVSWSYKREIK 3112
 QY 3226 PGDKYSLRQEGAMJELVVRNLRPODSGRY-----SCSFGDQTTSATLVTALPAQFTGKLR 3281
 Db 3113 PSREFRMTQFBDTYQLBIAEAYPEDEGTYTFVASNVSQVTSAILKLEAPEKIMYEKLE 3172
 QY 3282 NK-----EATGATATLRCELSKTAPE--EWRKSETLRDGRYCLR 3321
 Db 3173 EEIEMEVKAPILRRRLLEPLEVAVNHVAKTCEVETTPNVKFWYKAGREIYDQKYSIR 3232
 QY 3322 QDGAMCELOIRGLAMVDAEYSCVCGERTSAS-----LTIRPMPAHFGLRHCESIEGA 3377
 Db 3233 SSNYLSTLEIPRPQVDCGEYSCASKNQHSVSSSTAFLTVE-PPRFIKLDSRLVKQH 3291
 QY 3378 TAT-LRCEL--SKAAPVEMKRGESLADGRHSLRQDGAECLEIQGLAVADAGEYSC-- 3432
 Db 3292 DSTRYECKVGSPEIKVTYKGETEHPSEKYSNSFVDSVAVLEMHNLVSDESDGYSCEA 3351
 QY 3433 --VCGERTSATLVKALPAKFTGLENEEAVGATAMLMCELSKVAP--VEWRKGPENL 3488
 Db 3352 QNPAGSASTSTSLVKAPPA-FTKPPHPVQTLKGSVDHLECELOGTPPQISWYKDKREI 3410
 QY 3489 RDGRYILRQEGTCELEIQGLAMADAGEYLCVC-----GOERTSATLITRALPARIEDV 3544
 Db 3411 RSSKKYKMGSENYLASIHLNVTADYGEYHCKAVNDVGDSDGICSGVTLRA-PPTFVKKL 3469
 QY 3545 KQEARSGATVLOCELSNAAPVE--WRKG-SETLRDGRYSLRQDQTKCELOIRGLAMA 3601
 Db 3470 SDVTVVVGEIIELOAAVEGQPTSVLWLDKKGIIRESENLWISYSENVASLIGNAET 3529
 QY 3602 DTGYSVCVCGERTSAMLTVRALPIKFTETGLRNEEATGATVLRCELSQAPVEMWKGH 3661
 Db 3530 NAGKYIC----- 3536
 QY 3662 ETLRDGRHSLRQDGAECLEIQGLVAEDAGEYLCMGKERTSAMLTVRAMPSKTEGLR 3721

Db 3537 -----QIK-----NDAGFQECF-----AKLTV-LEPAVIVEKPG 3564
 QY 3722 NEEATEGTATLWCEL--SKAAPVEMKRGHETLRDGRHSLRQDGRSCELQIRGLAVUDA 3779
 Db 3565 PVKVTAGDSCTLECTVDGTPELTARMPXDNELSDTHKYKISFFNKVSGLKLNAGLEDS 3624
 QY 3780 GEYSC-----VCGERTSATLV--RALPARIEDVKNQEARREGATAVTQCELSKAAP--V 3831
 Db 3625 GEYTEVKNVSGKSSCTASLOVSDRIMPSPSTRKLKETYQGLGSSAVLECKVYSGPFLV 3684
 QY 3832 EWRKSETLRDGRYSLRQDGRCELOIHLISVADTGEYSC-----VCGERTSATLTVRA 3887
 Db 3685 SWFHDQBITSGDKYQATLTNTCSLKVNGLQESDMGTYSCTATNVAGSDCSAFLSVRE 3744
 QY 3888 POPVF--REPLQSLQAEESTATLQCELSBPTATVVMKGLQLOANGREPRELQCTA 3944
 Db 3745 P-PSFVKPPEPNVLSGENIITFSI--VKGSPPLEVWFRGSIELAPGHKCNITLQDSVA 3801
 QY 3945 ELVLQDLQREDTGEYTCGSOATSATLV--TAAPVRFRLRELOHQEVDDEGTAHLCC 4001
 Db 3802 EELFDVQPLQSGDYTCQVSNAGKISCTTHLFVKEPAKFMKVDLSVEKG--KNLILE 3859
 QY 4002 LSRAGA--SVEMWKGSLQFPCKAYQWQDGAAILVRGVEQEDAGDYTC---DTGH 4054
 Db 3860 CYTGTPTSPITWKNQVILKHSEKCSITTTETSAILEIPNSKLEDDQOQYSCHIENDSQ 3919
 QY 4055 TQSMASLVRVPRPKFTRLQSLQSOETGDIARLCCQLSDAESGAVVQWLKGVHLAGPK 4114
 Db 3920 DNCHCAITLSP-PFVFTPLBPVQTVGDSASLOQVAGTPE-MIVSWYKGDTKLRGTAT 3977
 QY 4115 YEMRQCATRELLIHOEAKDTGEYACV--TGQKTAAAL-----RVTEPVTIVRGLV 4166
 Db 3978 VMHFPKQVATLVPSQVSDSDSGEYICKVENTGEBATSSSLTVOERKLPPSFT--RKL 4035
 QY 4167 DAEVTADEVDVFCSEVSRAGATGVQWCLQGLPLOSNEVTEVAVRDGRIHT-----LR 4218
 Db 4036 DVHETVGLPVTFDCGIAGSEPIEVSFKDNVRV-----EDYNVHTSFDINVAIQ 4086
 QY 4219 L-----KGVTPEDAGTVSPHLGNHASSAQLTVRAREVITL--EPLQDVQLSEGQASQCR 4272
 Db 4087 ILKTDKSLMGVYTCASNAIGTASSGKVLTEGKTPFPFDTPTIPVDGIIGESADFECH 4146
 QY 4273 LSRASGQEARWA 4284
 Db 4147 ISGTQPIRVTTWA 4158

RESULT 4

T29757

protein UNC-89 - Caenorhabditis elegans

C.Species: Caenorhabditis elegans

C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999

C.Accession: T29757

R.Du, Z.; Le, T.T.; Wilson, R.

submitted to the EMBL Data Library, May 1997

A.Description: The sequence of C. elegans cosmid C09D1.

A.Reference number: Z20679

A.Accession: T29757

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-6642 <DUZ>

A.Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN000019; CESP:unc-89

A.Experimental source: strain Bristol N2; clone C09D1

C.Genetics:

A.Gene: CESP:unc-89

A.Map position: 1

A.Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6

/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 6.0%; Score 2489; DB 2; Length 6642;

Best Local Similarity 20.8%; Pred. No. 1.3e-65;

Matches 1520; Conservative 972; Mismatches 2737; Indels 2064; Gaps 304;

Db 2284 E--SVKPSAIVTGKPMFNVVYLNKKLIQSEEVKVKVHETGKTSIRIOKPLMEHN--- 2338
 QY 2052 ADRGFFGCTP-----DKTOA-KLTVM--ROVLRVGLQAVEARE-QGT 2093
 Db 2339 ---GTRVEAENVSGKVQATQALQVKKTEVPKFTTNWDRQVK-----EGEDVKFT 2387
 QY 2094 ATMEVQLSHADVDSWTRDGLRFQOGFTCHLAVGPMHTITLSGLRPEDSGLMVFRAG- 2152
 Db 2388 ANVE---GYPEPSVAWTLNGSPFVKPHENITVDKDGSEHTIEISAVTEQAGEISCEATNP 2444
 QY 2153 VHTSARLY-----VTELPVSFSRPLQDVVTEKEKVTLECELS--RPNVDVRLWKDGVE 2204
 Db 2445 VGSKRQVQLAVKKVGDP--TFAKLEDRLLITEGELTLMDAKLVNPKPKITWLKDGVE 2503
 QY 2205 LRA-GKTMAIAAQACRSLTIYRCFADQGVVCDADH-----AQSASVKVQGRTYTLIY 2259
 Db 2504 ITSDGHYKIVEEEDGSLKLSILOTKLEDKGRITIKAESFQVABCSASLGV-----VK 2556
 QY 2260 RRVLAEDAGEIQFVAENASRAQLRVKELPVTLPRLDKIAMEKHRGVLECOV---SRA 2316
 Db 2557 GRPKAKPA-----FQSDIA-----FINLT-----EGDTLECKLLITGDP 2590
 QY 2317 SAQVRWFKGSQELQPKYELV--SDGLYRKLIIISDVHADEDETYTC-----DAGDVKTSAQ 2371
 Db 2591 TPFVKWYIGTQLVCATEDTEISNANGVY-TMKIHGVTADMTGKIKCVAYNKADEVSTEG- 2648
 QY 2372 FFVEEQSITIVRGLQDVTVMEPAWPE--CETSI-----PSVRPPK 2411
 Db 2649 -----PUKVAPIVBEFETSLUCDRCREGDTLKLRAVLLGEPBPV---S 2690
 QY 2412 WLLGKTVLQAGNVGLEQEGVHRLMLRRTCSNWTGPVHFV-----GKSRSSARLVVSDI 2467
 Db 2691 WYANGKLEESONIKIHSEKGYTVTIKIDTCDYSGQVCEAINYEGKATSEATL----- 2745
 QY 2468 PVULTRPLEP-----KTG-RELQSVLSCDPRPAPKAVQWYKDDTPLSPSEKPK 2515
 Db 2746 -LVLPGRPPPLFELWLSNVRARTGKVHKVVFVGDPKP---SLTWYINKEILNSDLYT 2801
 QY 2516 MSLEGQMAELRILRMP-----ADAGVYRQCA-----GSAHSSSTVTVVEARE 2557
 Db 2802 IVTDKTTSLTNSFNPVHVGEIICKAENDAGEVSCANWITYSDMPSESSESAQAE 2861
 QY 2558 -----VTVTGP-----LQABATEBGWASFSCELSHEDB---VE 2589
 Db 2862 FVGDDLTEDESLREEMHRTFPPMAPKFIITKIDTKAKGHSVAFECV--PDTKGVCCK 2919
 QY 2590 WSLNGMPLYNDSFHEISHKGRHT-----LVKSIORADAG-----IVRASSLKV 2634
 Db 2920 WLKDGKEI-----ELIARIVQTRTGPBGHITQELVLDNVTPEBAGKYTCIIVENTAGKD 2973
 QY 2635 STSARLEY-----RVKPVVFKALDLSABERGTLALQCEV--SDPEAHVVVRKDG- 2684
 Db 2974 TCEATLVIESLEKKSEKKAPEFVALQDTKTKTSEKVLCKVIGBPKPKVSWLHDNVS 3033
 QY 2685 -QLGSPDKYDPLHTA-----GTRGLVHVDVSPEDAGLYTCHV-----GSEETRAVRVH 2732
 Db 3034 REKNPSEKITQESITIVESVEGVERVITTSSELSHOGKYTCIAENTEGTSKTEAFLTVQ 3093
 QY 2733 DLHVGITKRLKTMELVEGESCFEVLSHESASDPAMWTVGGKT-----VGSSSRF----- 2783
 Db 3094 GEAPVFTKELQNKELSIGELVLSCSVKSGPQPHVDYFSPSETTKVETKITSSRIAIEH 3153
 QY 2784 -QATROGKTYI-LVVRAAAPSDAGEVVPVSRGHTSKASLIVRERPAAIKPLEDQWAPG 2841
 Db 3154 DQTNHRMWVISQITKEDIIVSYKAIATNSIGTATSTSKITTKVEAPVFBQGLKKTSVKEK 3213
 QY 2842 EDVBLRCELBRAGTPFWHLKDRKAIRKSQYDVV--CEGTWMLVIRGASLKDAGEYTCE 2899
 Db 3214 EEIKWVKVGGASPDVVEFKDDKPVSDGNHMKONPETGVTFLVVKQAATTDAGKYTAK 3273
 QY 2900 VEASKSTASLHVEKAN-----CFTEEL--TNLQVEEKGTAFTCTKTEHPAATVTVRKG 2951
 Db 3274 ASNPAGTAESSAAEAVTQSLEKPTFVRELVTTEVKINETATLSVTVKGV--PDPSVEWLKD 3332

QY 2952 LLELRASGKHQPSQ-EGJ-TLRLTISALEKADSDTYTC-----DIQQAQSRALQLVQGRV 3005
 Db 3333 GQPVQTDSSHVIATKEVSGSYITIKDARLEDSGKYACRATNPAGEAKTEANFAVVKNLV 3392
 QY 3006 --HIILEDVVDVQEGSSATFRCRISPANYBPVFWFLDKTPLHANELNEIDAQFG-GXHV 3062
 Db 3393 PPEFEKLSPLVEKESSTLISVKVGPPEPSVEMFKDDTTPISIDNVHVIQOTAVGSFS 3452
 QY 3063 LTLROLALKDSOTIYF-----EAGQRAAALRVTEK--PSVFSRELTDAITTEGEOJTL 3115
 Db 3453 LTINDARQDVG-IYSCRARNEAGEALTANFGIIRDSIPPEFTQKLRPLEVREOETL 3511
 QY 3116 VCETSTCDIP-MCWTKDKGTILRGSARCQLSH-----EGHRAQLLITGATLQDSRYKC 3167
 Db 3512 KVTVTGTPVNVFVKDDKPI-----NIDNSHIFAKDESGHHT-LTIKQARGEDVGVVTC 3566
 QY 3168 ----EAGACACSSIVRVHAR---PVRFOALKDLVLEGGGAATLRCVLSSVAAP-VKWCY 3219
 Db 3567 KATNEAGEAKTANMAVQEEIEAPL-FVOGLKPYEVEGKPAELVVRVGEKPEPEVKFK 3625
 QY 3220 GNVLRPDGKYSLSROEG--AMLELVRLNRPDQSGRYSC-----SFGQOTTSATL----- 3267
 Db 3626 DGVPIADNQHVIKKGNGSGHTLVIKDTNNADFKYTCQATNKAGKDETVGELKIPKYS 3685
 QY 3268 ---TVTALPAQFQIGLKRKEATGATATLCELSKTA--PVEWRKGSSETLRDGDYCLR 3321
 Db 3686 FEKQTAEBVKPLFIEPLKETFAVSGDTVWLECKYNKESHPOIKFFKNDQPVIEIQHMLE 3745
 QY 3322 --QDGMCELQIRGLAMVDAABYSC---VCGERTSASLTI----- 3357
 Db 3746 VLEDGNI-KLATQNAKEDVGAYRCEAVNVAGKANTNADLKIQFAAKVEEHVTDESGOLE 3804
 QY 3358 -----RMPBAHFTGLRKHQSIIEGATATLCELSSKAAP---VEMRK 3395
 Db 3805 EIQOPEVTDGTSKTDTRGAP-EFVELLRSCVTTEKQALLKCV-KGEPRKIKWTK 3862
 QY 3396 -GRESLRDGRHSRQDGAVALCEIQICGLAVADAGEYSCVCGERTSA-----TLTVKAL 3448
 Db 3863 EGKEVENSARVRAEHKDDGTULTLTFDNTVQADAGEYRCEAENEYGSATWTEGPIIVTLEGA 3922
 QY 3449 P-----AKFTEGLRNEEAVEGATAMLWCELS-KVAP-VEMRKGPENLRDGDYILR--Q 3498
 Db 3923 PKIDGEAPDFLOPKPAVTVVGETAVLEGKISGPKSPKSVKYNKEELKPSDRVKIENLD 3982
 QY 3499 EGTRCELOICGLAVADAGEYLCVCGQE---RTSATLTIRALPAR---FTEDEVKNOEAR 3550
 Db 3983 DGTQ-RLVTWNAKDDMDXEYRCEASNEFGDVMSDVTLTVKE-PAQVAPGFFKELSAIQVK 4040
 QY 3551 EGATVLOCELNSAAP-VEMRKGETLRDGR--YSLRQDQTKCELQIRGLAMADTGEY- 3606
 Db 4041 ETETAFKFGSGTKPDVKFKDGTPLKEDKRHFESTDDGTQ-RLVIEDSKTDQGNR 4099
 QY 3607 ---SCVCCOERTSAMLTVRALP---IKFTEGLRNEEAETEGATVLRCEL-SKMAPVEMWK 3659
 Db 4100 LEVSNDAVANSKVPITV--VPSETLXIKKGLTDVNTVQTKILLSVEVEGKPKTVKNYK 4157
 QY 3660 GHETLRDGRHSRQ-DGARCELOIRGLVAEDAGEYLCMCQKE---RTSAMLTVR----- 3710
 Db 4158 GTETVTSQTTKIVQVTESEYKLEIEGAEMSDGTGAYRVLTSDTSFVSSESATVTVTKAAE 4217
 QY 3711 --AMPSKEIEGLRNEEAETEGATLWCEL-SKAAPVEMRKGHETLRDGRHSRQDGRSC 3767
 Db 4218 KISLPS-FKGLAQGVSPKGPVLVEIEEGKPDVKYKNGDEIKGKVEDL--GNGKY 4274
 QY 3768 ELQIRGLAVVDAGYSKV---CQOERTSATLTVRALPARFIEDVKVQNEAREGATVLOQ 3823
 Db 4275 RLTIIPDFQKDVGEYSVTAANEAGEIESKAKVNSAKP-EIVSGLVPTTVKQGETATFNV 4333
 QY 3824 EL-SKAAPEVRKGSSETL-----RGDRYSLRQDQTRCELOIHGLSVADTGYSCV-- 3873
 Db 4334 KVGPKVGVKVKYKNGKEIPDKAKTDKNDGGSYS-----LEIPNAQVEDAADYKVVVS 4384

| | | | | |
|----|------|--|--|------|
| Qy | 3874 | -- | CCQERTSATLTVRAP-----QPVRBPLOSQAEBGSTATLOCEISEPBTATVWSK | 3922 |
| Db | 4385 | NDAGDADSSAALTVKIADGCKVKPKEIVSGLIPTTVKQOETATFNVKGPVKQVKWKV | 4444 | |
| Qy | 3925 | GGIQLQANGRPRLGCGTAELVLQLOREDTCEYTCTCGSQ-----TSATLTVTAAPVR | 3980 | |
| Db | 4445 | NGKEI--PNAKAKONGDG--SVSLIIPNAQLDDTDADYKVVVNSNDAGDADSSAALTVKI | PGIA | 4502 |
| Qy | 3981 | FLRELOHEVDEGGTAHLCCELSRAGASVEWRKSLQLPFCAKYQMVDQG--AAAE | LLVRG | 4039 |
| Db | 4503 | IVKGLDAEVPKGGKAVLQVETNNKPKEIKWYKNGKEITPSDKAQPGSDGNKPOLVIPD | 4562 | |
| Qy | 4040 | VEQEDAGDY---TCDTGH--QSWASISVRVP--RKFKTRLQSLQOETGDI--AR | CCQL | 4091 |
| Db | 4563 | AGDDAAEYKVVUTDEGNTADSSCALTVKLPKPEKI---IKGLDQVVVISGSP | IKLEI | 4619 |
| Qy | 4092 | SDAESGAVQWMLKEGVELHAGPYEMRSOGATRELLTHQLEA---KDTGEYACVTGGQKT | 4148 | |
| Db | 4620 | ETSGSPKTVKWKNGKELPGAAAKTIKIQKIDNNKYVLEIPSSVVEDTGDYKVEVANEAG | 4679 | |
| Qy | 4149 | AA--SURVT--EPEVTIVRGLVDAEVTADEDEVSECVSRAGATGVQWICLQGLPLQ | NEVT | 4205 |
| Db | 4680 | SANSSGKITVEPKITFLUKPKDOOSITEGENAEFSVETNKPRI--VKWYKNGQEI | KPNR | 4738 |
| Qy | 4206 | EVAVRGRIRHLKLGVTPEADGTVSPHLCNHA-----SSAQLTVRAPEV---TILE | PLQD | 4258 |
| Db | 4739 | IIEQKTDTKVQLVIKNAVRDADTYKIVLENTAGEAESQAULTVKKAKAGLCXIVK | GLE | 4798 |
| Qy | 4259 | VQLSEGDGAGFCQRLSRASQEARWALGGVPLQANEMNDITVEQGLHLHLTHKVTLEDA | 4318 | |
| Db | 4799 | QVVAKAKMVFVKI--QGEPEDEVRLRDANVI SAGANAIIEKIDDDTYRLLIIP | SADL | 4857 |
| Qy | 4319 | GTVSFHV---GTCSSBAQLKVTAKNTVRGLENVREALGEALFECOLSQBPVAAHTWL | 4374 | |
| Db | 4858 | GEYTVINESGKAKSDAKEVEDEKPEIVRGLENIETPEGDDVDVKVEVSAP--VRQVKWY | 4916 | |
| Qy | 4375 | LDPEVRTSNAEV-----VFFENGLRHLHLKLNLRPDQSCRVTFLAGDMV | 4420 | |
| Db | 4917 | KNDQEIKNPHLEAKKIGPKYELAINRAQLDDGADYKVLSN-----AAGDCD | 4965 | |
| Qy | 4421 | TSALFT--VRGWRLEILPLKNAAVRAGAQAARFTCLSEAVPVGSEAWYINGAAVQDDSD | 4479 | |
| Db | 4966 | SSAALTIVKPNVILKIVDGLKDVDEBPQVELKVKV--EGIP--KVIKWYKNGQELKPD | --- | 5020 |
| Qy | 4480 | WTVTADGSHQALLRSQPHHAGEVTFACRDVASARLTVLGLDPDPPEAEVVAAHSHTV | 4539 | |
| Db | 5021 | ----ADG-----FKPEEKPESC-----FSL | 5037 | |
| Qy | 4540 | TLSWAAPMSDGGGLCGYRVEVKEGATGQWRLCHELVPGECEVVDGLAPETRYFRVA | 4599 | |
| Db | 5038 | TIP--SSXKSDGG---AYR-----VVLGNDKGEV----- | 5062 | |
| Qy | 4600 | GPVGAGSEFVHLPOTVRLAEPKPVPOP--SAPESROVAAAGEVDSYLEVVAEA--GE | 4656 | |
| Db | 5063 | ----SGSVVH-----VKSAKSSEPTSGANFLSPDKTEVBEGDMLTLCQT | IAGE | 5114 |
| Qy | 4657 | HKGVERTQPGGRFV--VSGRQOQMLVTKGTAEADQGEYHCGLAGGSICPAAATFO | VALSP | 4715 |
| Db | 5115 | EKDGWVLQKDDRIITMRVALDGTATLRISAKSDIGQYRV-TAKNEAGSATSCKVT | VT-- | 5172 |
| Qy | 4716 | ASVDEAPQPSLP-----PEAAQEGDLHLLEALAKRMSREPTL-----D | 4756 | |
| Db | 5173 | ----EQGQPKPKFVIFLKTGAALPGDK----KEFNVKVRLPKPTLQWFLNGIP | KFDD | 5225 |
| Qy | 4757 | SISELPBEDGRSQRLPQAEAEVADPLSEGYSTADELARTGDADLSHTSSDDESRAG | PPSL | 4816 |
| Db | 5226 | RITLDDNADGNVCLTIRDVRE-----EDFQTLKCIK-----NENG | TD-- | 5266 |
| Qy | 4817 | VTYLLKAGRGCTSPSLAKSVGAPAPSVKPPQQQEPPLAAVRPLLDLSTKDLG | PSMD | 4876 |
| Db | 5267 | CEFOQGAGH-----DDGRDDLRVP----- | 5286 | |
| Qy | 4877 | VKIQAPAFKGYKVRKEMKQEGPMFSHTFGDTEAQVGDALRECVASAKADVARWLK | XDGV | 4936 |

| | | | |
|----|------|---|------|
| QY | 980 | VVFAKDQVAHSEVQAEAGANATLSC-EVAQAQAEVVMYKDGKGLSSSLKVHVFEAKGRRR | 1038 |
| DB | 795 | II---ESPHT-VRVNIERQVTLQCLAVGIPPEIEWQK-GNVLLATLNNPRVYTLQADGN | 848 |
| QY | 1039 | LVVQQAOKTADAGDYSCERG----QRVSFLHIT---EPOMFAKQSVHNEVQAEAGAS | 1091 |
| DB | 849 | LLITDAQIEDOGFTCIARNYQOQSGSTLMTGLVSPVL-----GHVPEPESQLIEGQD | 903 |
| QY | 1092 | AMLSCEV--AAQATEVTWYDKGKLSKSSKVGMEVKGCTRRLLVPAQK-ADAGEYSCEA | 1148 |
| DB | 904 | LTLSCVVLGTPKPSIWIWDXDPVEBGPITKIEGGSLRL---RGNPKDESKYTCIA | 960 |
| QY | 1149 | GGORVSFLHIT-----EPKGVFAKE-----QSVHNEVQAEAGYT | 1183 |
| DB | 961 | VSPAGNSTLHINVLIXKPEFVVKYPEGGIVFKPTISGMDEKHVAVVNSTHDLVDGEG-- | 1017 |
| QY | 1184 | AMLSCEVA-OPQTEVTWYKDGKLSKSSKVRMEVKGCTRRLLVVQOVQKADAGEYSCEA-- | 1240 |
| DB | 1018 | FAIPCVSGTTPPIITWYLDGRPITPNS--RDTVTADNTLIVRKADKSVGYVTCQATN | 1075 |
| QY | 1241 | ----GGORVSFLHITBPKAVFAKEQLVHNEVTEAGASATLSEV-AAQATEVTWYKDG | 1295 |
| DB | 1076 | SAGNECKTTRIMNT---PMISPGQSSFNWVDDL---FTIPCDVYKDPKPVITWLLDD | 1129 |
| QY | 1296 | KLSSSSSKVRIEAGCMRQ---LVVQAGADAGEYTC-----AGGQRLSFHLVDSEPKA | 1348 |
| DB | 1130 | KPTE-----GVNEDGSLTIPNVNEAHRGTFTCHAQNAAGNDRITVTLVTTTPT | 1180 |
| QY | 1349 | VFAKEQLAHRKVOAEAGAIATLSEVAQAQTEV-TWYKDGKLSSS---SKVRMEAVGCT | 1404 |
| DB | 1181 | INAEHQ--EKIALQNDI-VLECPAKALPPVRLWYEGEKIDSQLIPTHIREDGA-- | 1233 |
| QY | 1405 | RLVWQQAQADATGEYSCE-----AGGQRLSFSLDVAE-PKVFAKQGPVREVQAAQAGAS | 1459 |
| DB | 1234 | --LVLQNVKLNTGTFVFCQVSNLAGEDSLTYTLVHEKPII---SEVPGVVDVVKGFT | 1287 |
| QY | 1460 | TTLSEVAQAQTEW--WYKDGKLSF-SKVRMEAVGCTRRLLVWQAGADAGEYSC-- | 1514 |
| DB | 1288 | IEPCR-ATGVPEVIRTWNGIDLKWDEKFSVDNLGLR---IYEADKND-GNYNCVV | 1343 |
| QY | 1515 | --EAGSORLSFHLHABPKAVFAKEQPASREVQAEAGTSATLSEV-AAQATEVTWYKDG | 1571 |
| DB | 1344 | TNEAGTSQMTTHVDVQEPPIILPSTQNTAV--VGDRVELKCYVEASPASVTFWFRG | 1400 |
| QY | 1572 | KLSSSSK-VRMEAVGCTRRLLVQEAQADAGEYSCKAGDQ--RLSFHLHV---ABPKVV | 1625 |
| DB | 1401 | IAIGDTKGYVVEDG--TLVQASVEDATYITCKASNPAGKAEANLQVTIAGPDI-1456 | |
| QY | 1626 | FAKEQPAHREVQAEAGASATLSEV-AAQATEVTWYDKGKLSKSSSKVRVEAVGCTRRLV | 1684 |
| DB | 1457 | --KDPDVVTQESIKESHPPFLYCPVFSNPLPQISWYLNDRPL-IDDKTSWKTSDDKRKLH | 1513 |
| QY | 1685 | VOQAGADAGEYSC-----EAGQRLSFRLHVAELEPOISRPCREPLVVKHEHDIILTA | 1740 |
| DB | 1514 | VFAKAITDSGVYKCVARNAAEGESKSPQVEV--IVPLNDBSKYKVKVFAKEGEEVTLGC | 1571 |
| QY | 1741 | TLATPSAATVTLKDGVEIRRSKEHTAS-QGDPTHTLTVHGAQVLDSAIYSCRVAGQD | 1799 |
| DB | 1572 | PVSGFPVPQINWVVDGTVVEBPKKYKATLNDGLTLHFDSVSVKQBGNYHCVAQSKGNI | 1631 |
| QY | 1800 | FPVOVEVAAKFCRLBEPVCGE-----LGGTVTLACEI---SPACAEVVRWCGNTQPR | 1849 |
| DB | 1632 | LDIDVE-----LSVLAVPIVGEDDNLEVLFGKOTLSLSCDQTESDDDKTTTFWWSINSESD | 1686 |
| QY | 1850 | VGKRFQWAEQVPSRLTVLGLRAEDAGEYVCESDHTSQAOLTVSFV-----1895 | |
| DB | 1687 | RPDNNVQPSDG--HELITDAPENNGKMYCVRVNSAGKAERTLTLDLBPPVFVEPVE | 1744 |
| QY | 1996 | -----PRVVKFMSG-----LS | 1906 |
| DB | 1745 | ANQKLIIGNPITLOQV'TGNPKPTVWIKIDGNDVDKSWLPDESLSLRIEKLTKSAQIS | 1804 |

| | | | |
|----|------|---|------|
| Qy | 1907 | TVVAEBGREA-----TFQCVVSPSDVAVVVPRDGA | 1936 |
| Db | 1805 | CTAENKAGTASRDFFIQIAAAPTAKNEGQOETIFRESEITILDCPVSLGDFQITWKQGL | 1864 |
| Qy | 1937 | LLQPSB-----KPAISOGASHSITISDLVLEDAGQITVAEAGASSAALRVREAPV | 1988 |
| Db | 1865 | PLTENDAIPLDNRTRLTLINARDHEDIYTCVANNTAGQVKDFVV-----VQV | 1914 |
| Qy | 1989 | LFKKLEPQTV-----ERSVTTLEVELTRPWPRLWRTRNATALAGKNVBIHAEGARHRL | 2044 |
| Db | 1915 | LPKIKNAVVTLEINEGEIIILTCDAE-GNPTPTAKWDFNOGDL-PKEAVFVN---NNHTV | 1969 |
| Qy | 2045 | VLHNVGADRGFGCEPDDKTKOAKITVEMR---OVLVRGLQAVAREOGTATMEVOLS | 2101 |
| Db | 1970 | VNNVTKYHTGVYKCVATNKVGQAVKTIINVHVTKPRFESGLTESELVNLFRSITLECD | 2029 |
| Qy | 2102 | HADVDG---SWTRDGLRFQOQPTCHLAVRGPMTLT-----LSGLRPEDSG---LMVFK | 2149 |
| Db | 2030 | VDDAIGVGISWTVNGKPF-----LAETDGVQTLAGGRFLHIVSAKTDHGSYACTVTN | 2082 |
| Qy | 2150 | AEGVHTSARLVVTLPVPSRPLCDVVTTEKEKVITLECELS-RPNVDVWKLGDVLRAG | 2208 |
| Db | 2083 | EAGVATKTFNLFVOVPPTIIVNEGGEYTVIENNSLVLPCVEVTGKPNVVTWKDGRPVGDL | 2142 |
| Qy | 2209 | KTMIAAQAQACRSLTIYRCEFAQGVVYCDAHDAQSSASVKVQGRVTYLIYRVLAEADAG | 2268 |
| Db | 2143 | KSVQVLESG-QQFKIVHABIAHKGSYICVAKNDVGTAEISPD---VDIITRPMI----- | 2192 |
| Qy | 2269 | EIQVVAENASRAQLRVKELPVTILVR-----PLRDKTAMXKRGVULCQVSRASQVRWFK | 2324 |
| Db | 2193 | -----QKGIKNI-VTAIKGALFPKCPIDDDKN-----FKGQIIMLR | 2328 |
| Qy | 2325 | GSQ---ELQPGKYELVSDGLRVKLIISDVHAEDEDTYTC---DAGDKVTSQAPFVEE | 2376 |
| Db | 2229 | NYQPIDLEADARITRLSND---RLTLINVTENDSGQYSCRVNQDAGE--NBFDF--- | 2279 |
| Qy | 2377 | QSITIVRGLQDVVMEPAPAWFCECTSIPSVRPPKMLLGTVLVQAGNVGLEQEGTVHRL | 2436 |
| Db | 2280 | -----KATVLVP----- | 2286 |
| Qy | 2437 | MLRITCTMTCPVHFTYVKGSRSSARLVVSDIPVVLITPLEPKTGRELQSVVLSCDFPAP | 2496 |
| Db | 2287 | -----FTIIMLDKDN-----KTAVEHSTVTLSCDPATGKP | 2316 |
| Qy | 2497 | KA-VQWYKDDTPLSPSEKFKMSLEGOM-AELRILRLMPADAGVYRCQAGSAHSSTEVTV | 2553 |
| Db | 2317 | EPDITWFKDEAIIHENIADIIPNGELUNGQKAITRIKEDGACKYTCEADNSAGSVEQDV | 2376 |
| Qy | 2554 | EAREVTV-----TQPLQOABATGEWASFSCEL-SHEDEEVEVSLNGMPLYNDSPFEISH | 2607 |
| Db | 2377 | NWNVITPIKXKDGPISDYESQQNERVVIISCPVYARPPAKITWLKAGKPLQSDKPFKXTSA | 2436 |
| Qy | 2608 | KGRHTIIVLKSIORADAGIVRASLSKVYTSARLAEVRVKPVVFLKALDLSAEERGTL--- | 2664 |
| Db | 2437 | NGQK--LYLFKLRETDSSKYTCIATNEAGTDKDFVKVSLV-APSPDEPNIVRRTVNSG | 2493 |
| Qy | 2665 | ---ALOCEV-SDPEAHVVRKDGVLQSPSKYDPLHTAGTRGLVHVDVSPEDAGLYTC-- | 2718 |
| Db | 2494 | NPSTLHCPAKGSPSFTIWLKDGNAIEPNDRYVFF-DAG-ROLQIKSTEGSDQGRYTCIA | 2551 |
| Qy | 2719 | --HYGSETRARVRVHDLHVGLTKRLTKMEVLGECSFECVLSHESASDPAMW----- | 2771 |
| Db | 2552 | TNSVGSDDLENTLEVIIIPVWDGERRRAVAVIGFSESELF--DSNSTGVYDVWEQKDGLT | 2609 |
| Qy | 2772 | VGGKTVGSSSRFQATROGRKYIILVYREAAPSDAGEVVVFSVRGLTSKA----SLIVRBPRA | 2827 |
| Db | 2610 | INQDTLAGDSFIQIPSSGKKMSFL--SARKSDSGRYTCIVRNPAGEARKLDFPAVNDPPS | 2667 |
| Qy | 2828 | AI---IKPLEDQWAPGEDVELRCELSRAGTTP-VHWLKDRAIR-KSQKYDVVCEGITWAML | 2883 |
| Db | 2668 | ISDELSSANIQITIIYFYFVEINCVVSGSPHPKYVWLFDDKLELPEDPSAAYELTNNGETLKI | 2727 |
| Qy | 2884 | VIRGASLKADAGEYTCVEAS-----KSTASLHVEBKANCFTBELTNLQVKEKGT | 2933 |

| | | | | |
|--|----|---|--|------|
| | QY | 4938 | LTDGR-----HHHIDQLGDTCSLLIAGLDRADAGCYCQVSNKFGQVTHSACVVVGSES | 4993 |
| | Dd | 4610 | IEDSKFELAYSHLKVTLNST-----SDSGEYTCMAQNSVGSSTWAFINVD---- | 4655 |
| | QY | 4994 | EAESSGGGLDDAFRAARRHLRFLRTKSPAEVSDEBELFILSADGPAPREPADWTQTYRE | 5053 |
| | Dd | 4656 | -----NNILFTPKPS-----SNQKNVAVITCYER | 4679 |
| | QY | 5054 | DSHFICIRFEALTEARQAVTRPQEV-----FATLG--IGVEIKLVEQGPRVRVMCIKETP | 5107 |
| | Dd | 4680 | NQAY-----SRGLTWEYNGVPMPKNLAGIHFWNNGSLVILDTSSLKEGDLELYTKVNRNR | 4735 |
| | QY | 5108 | APVWP-----BPLSLTSDAAPVLFTLONQEVDQGYPVSFVCVTTGOMPMS-VRWFK | 5161 |
| | Dd | 4736 | RHSIFHLTSAFEVGEVKTIKV-----EYNGDSVVLCEVTSPDLATHVVMWK | 4785 |
| | QY | 5162 | -DGKLEDDDHVMINEDQQGHOLIITAVPADMGVYRCIAENSMGVSS"KAELRV----- | 5216 |
| | Dd | 4786 | NQKMLDADAIVL-----PNNSLVLLNVEXYDGVYKCVASINSIGAKAFDDTLQNVEGD | 4840 |
| | QY | 5217 | DLUTSDYTAA-----DATESSYSFAOYLSSRSRGEGTESTETDGGQLPQVVEELRDLOVA | 5272 |
| | Dd | 4841 | FPLTGFGGSGGNIDDSSNAGSSRRRAYKKENEDASTTTTTTSPITTTTETPLATTII | 4900 |
| | QY | 5273 | PGTRIACKOLKVKGYP | 5288 |
| | Dd | 4901 | P-----ALITLPKQIP | 4912 |
| | | RESULT 6 | | |
| | | T20992 | hypothetical protein F15G9_4a - Caenorhabditis elegans | |
| | | C:Species: | Caenorhabditis elegans | |
| | | C>Date: | 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000 | |
| | | C:Accession: | T20992; T24733 | |
| | | R:Sulston, J. | submitted to the EMBL Data Library, December 1994 | |
| | | A:Reference number: | Z19355 | |
| | | A:Accession: | T20992 | |
| | | A>Status: | Preliminary; translated from GB/EMBL/DDBJ | |
| | | A:Molecule type: | DNA | |
| | | A:Residues: | 1-5175 <W1> | |
| | | A:Cross-references: | EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GNC0028; CESP:F15G9_4a | |
| | | A:Experimental source: | clone F15G9 | |
| | | R:Kershaw, J. | submitted to the EMBL Data Library, December 1994 | |
| | | A:Reference number: | Z19929 | |
| | | A:Accession: | T24733 | |
| | | A>Status: | Preliminary; translated from GB/EMBL/DDBJ | |
| | | A:Molecule type: | DNA | |
| | | A:Residues: | 1-5175 <W12> | |
| | | A:Cross-references: | EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GNC0028; CESP:F15G9_4a | |
| | | A:Experimental source: | clone T09B9 | |
| | | C:Genetics: | | |
| | | A:Gene: | CESP:F15G9_4a | |
| | | A:Map position: | X | |
| | | A:Introns: | 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3; 1251/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1 | |
| | | 1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 5077/1 | | |
| | | Query Match | 5.0%; Score 2043.5; DB 2; Length 5175; | |
| | | Best Local Similarity | 21.4%; Pred. No. 1.3e-52; | |
| | | Matches 1186; Conservative 773; Mismatches 2265; Indels 1309; Gaps 254; | | |
| | QY | 286 | VVEDAQ-----ENFVLKILFCQKS-----DRG-----LYTCTASNLYGQRY-- | 321 |
| | Dd | 136 | VFTDARKDYHLEDEVINTIQEKSVVFMVMTGDCGNETHPGFRTRYEKIAAASFQVPHL | 195 |
| | QY | 322 | -----SSVLVWVREPA-----VFFKGRLODLEVR-----EKES | 349 |
| | Dd | 196 | EKSDVSTVLEYVRHAVKQKHLMYEARERGCTVSRNPVDKHLSELTSISLSDGKDDSDN | 255 |

Db 1181 INAENQ---EKIALQNDI-VLECPAKALPPVRLWYEGEKIDSLIPHITREDGA--- 1233
QY 1405 RRLVQQAQADTGEYSCE---AGQRLSFLDVAE-PKVYFAKEQVHREVOQAQAS 1459
Db 1234 --LVQNKLKENTGVFCQVSNLAGEDSLSYTLTVHEKPKII---SEVPQVVDVVKGFT 1287
QY 1460 TTLSCEVAQAQTEVM--WKDKGKLSF--SKVRVEAVGCTRLVQVQAGQADAGEVSC-- 1514
Db 1288 IEIPCR-ATGVEPVRTWNGKIDILKMDKKFSVDNLGTUR---IYEAQNDIGNVCVV 1343
QY 1515 --EAGSRLSFLHVAEPKAVFAKEQBPASREVQAEAGTSATLSCEV-AQAQTEVTWYKDG 1571
Db 1344 TNEAGTSQMTTHVDVQEPPIILPSTQNTNTAV---VGRVELKCYVEASPASVTFWFRG 1400
QY 1572 KKLSSSK-VRMEAVGCTRLVQVQAGQADAGEVSCAGDQ--RLSFLHV---ABPKVV 1625
Db 1401 IAGTDTKGYVVEDSG---TJVIQASVEDATIYTCXASNAGKAEANLQVTVIASPDI- 1456
QY 1626 FAKEQPAHREVOAAGASATLSCEV-AQAQTEVTWYKDGKLSKSSKVRVEAVGCTRLV 1684
Db 1457 --KOPDVVTQESIKESHPSLYCPVFSNPLQISWYLNKPL-IDDKTSWKTSDDKKHL 1513
QY 1685 VQQAQADAGEVSC---EAGQRLSFLHVAELEPOISERPCHREPLVVEHEDIIITA 1740
Db 1514 VFKAITDSGVYKCVARNAAGEGSKSFQEV--IVPLNLDSESKYKKKVFAGEEVEVTLGC 1571
QY 1741 TLATPSAATVWLKDGVEIRRSKHEHAS-QGDPTLTLVHAQVLDLSAIVSCRVAGBQD 1799
Db 1572 PVSGFPVQINNVVDGTVFEGKYKATLSNDGLTLHFDPSVSVKQEGNTHCVAQSKGNI 1631
QY 1800 FPVQVEEVAKFCRLLEPVCEB-----LGGVTTLACEL---SPACAEVVMRGNTQPR 1849
Db 1632 LDIDVE-----LSVLAVPIVGEDDNLVFLGKOISLSCDLQTESDDKTTFVMSINGSSED 1686
QY 1850 VGKRFQVMAEVPVRLVGLRAEDAGEVYCESDDHTSALQTVSV----- 1895
Db 1687 RPDNVQIPSDG--HRLYITDAKPNNGKXCRVNTSAGKAERTLTLVDLEPPVPFVPE 1744
QY 1896 -----PRVXFMFG----- 1906
Db 1745 ANOKLIGNPPILOQVGNPKPTVIWKIDGNDVDKSWLFDSELSLLRIEKLTKSAQIS 1804
QY 1907 TVVAEGGEA-----TFQCVSPSVAVVVRDGA 1936
Db 1805 CTAEKAGTASRDPFIQNIAPTFKNEGDOETIFRESEITLDCPVSLGDFQITWKKQL 1864
QY 1937 LLOPSE-----KFAISQSGASHSLTISDLVEDAGQITVBAEGASSAALRVREAPV 1988
Db 1865 PLTENDAIFLTDNLTLNANRDHEDIYICVANNTAGQVSKDFV-----VQV 1914
QY 1989 LFKKKLEBPQVE-----ERSSVTELEVTRWPPELWTRNATALAPGNVBIHAGARHL 2044
Db 1915 LPKIKNAVVTLEINEGEEIITCDAE--GNPTAKWDFNQGD-L-PKEAVFN---NNHTV 1969
QY 2045 VLHNVGFADRGFCECPDDTKQAKLTVMER---QVRLVRGLQAVAREQGTATMEVOLS 2101
Db 1970 VVNVTKHTGVYKYATNKVQAVKTVINHVRTKPRFESGLTESELTVNLTSITLECD 2029
QY 2102 HADVDG---SWTRDGLRFOQGFCHLAVRGPMTLT-----LSGLRPEDSG---LMVFX 2149
Db 2030 VDDAIGVIGISWVNGKPF-----LAETDGVQTLAGRFLHIVSAKTDHGSYACTVTIN 2082
QY 2150 AEGVHTSARLVVTELPVSFRPLEDDVVTTEKEKVTLECELS-RPNVDVWLKDGVELRAG 2206
Db 2083 BAGVATKTNLFVQVPTTIVNEGEYTVIENSLVLPCEVIGKPNPVVWTKDRPVGDL 2142
QY 2209 KTMIAAQAAGACRSITTYRCEFPADQGVVCDADHAQSSASVKVQGRYTVTLIRVLAEDAG 2268
Db 2143 KSVQVLSG--QQFKIVHAETAHKGSYICMAKNDVGTAESFD---VDIITRMI----- 2192
QY 2269 EIQFVAENAESRAQLRVKELPVTILVR---PLRDKIAMEKHGVLBCQVSRASAOVRWF 2324

Db 2193 -----QKGKINI-VTAIKGGALPKFCPIDDDKN-----FKGQIIWL 2228
QY 2325 GSO-----ELQOPKVELVSDGLYRKLIIISDVHAEDEDTYTC-----DAGDVTSQAQFFVEE 2376
Db 2229 NYQPIDLEAEDARIIRLND---RELITINVTENDEGOYSCRVKNDAGE--NSDFD----- 2279
QY 2377 QSITIVRGLODVTVMEPAPAFECETSIPIVRPPKWLKGTVLQAGGNVGLQEQTVHRL 2436
Db 2280 -----KATLVP----- 2286
QY 2437 MLRRCTCTMTGVPVHTFVTKSRSSARLVVSDIPVWLTRPLEPKTGRELSVVLSCDRPAP 2496
Db 2287 -----PTIIMLDKDN-----KTAVEHSTVTLSCPATGK 2316
QY 2497 KA-VQVYKDDTFLSPSEKFKMSLEGOM--AEIRILRLMPADAGVYRCOAGSAHSTEVTV 2553
Db 2317 EBITWFKDGEAIHENTADIIPNGELNGNQLKTRIKEGDAGKTYTCADNAGSVEODV 2376
QY 2554 EAREVTV-----TGLODAEATEEGWASFCELS-SHEDEEVEWSLNGMPLYNDSFHEISH 2607
Db 2377 NVNVTIIPKIEKDGIPSDYESQONERVVISCPVVARPPAKITWLKAGKPLQSDKFKVTS 2436
QY 2608 KGRHTVLKSTORADAGIVRASSLKVSTARSARLEVRVKPVPLKALDLSAERSTL--- 2664
Db 2437 NGQK--LYLFLKRETDSSKYTCIATNEAGTKDRDFKVSMLV-APSFDFPNIVRITVNSG 2493
QY 2665 ---ALQCEV-SPEAHVVRKDXGVLQGPSDKYDFLHTAGTGLVVDVHVSPEADAGLYTC-- 2718
Db 2494 NPSTLHCPAKGSPSTITLWXDNALPNDRYVVF-DAG-RQLQISKTEGSDQGRYICIA 2551
QY 2719 --HYGSEETRVRVVDHVLHVGITKRLKTMVELEBESCFEVLCSHESASDDPMW---T 2771
Db 2552 TNSVSGDDELTLEVIIIPVTDGERREAVAVIEGFSSELF--DSNSTGVDVVEKQDGLT 2609
QY 2772 VGGKTVGSSSRFOATROGRKYLIVVREAAPSDAGEVWFSVRGLTSKA---SLIVRBRPA 2827
Db 2610 INQDTLRGDSFIQTPSSGKKMSFL--SARKSDSGRYTCIVRNPAAGEARKLDFAVNDPPS 2667
QY 2828 AI--IKPLEDQWVAGEDVELRCELSRAGTP-VHMLKDKAIR-KSQKYDVVCEGTWAML 2883
Db 2668 ISDELSSANIQTIVPYVPEINCVVSGSPHPKVWLFDDKPLEPDSAAVELTNNGETLKI 2727
QY 2884 VIRGASLKDAGEVCEVAS-----KSTASLHVEEKANCFTBELTNLQVEEKGTA 2933
Db 2728 V--RSQVHEAGTYTCEAQNNGVKARKDFLVRVTPPH-----FEKEREVVARVGDTM 2778
QY 2934 VFTCKTEH--PAATVTW-----RKGLLELRASGKHQPSQEGTLRLITISALEKASDST 2984
Db 2779 LLTCAESSVPLSSVYVWHAHDESQNGVI---TSKYAANEK---TLNVTNIQLDDEGF 2830
QY 2985 YTCDIGQASQAQLVQORRWHIIELEDVDVQV-----GSSATFRCPISPNYEPVHW 3038
Db 2831 YVC---TAVNEAGTKKFKFLIVTETPYFLDQOKLYPIILGKRLTLDG--SATCTPPTI 2885
QY 3039 FLDKTFLHANELNIDAQPGGYHVHLLRQLAKDQSGITYFE--AGDORASAAALRVTEKPS 3096
Db 2886 LFMKDKGRINSEDEVDI-IGSTLVINDPQKEVEGYTCIAENKAGRSEKMMVVELLPK 2944
QY 3097 VPSRELTAATTTEGEDTLVCE-TSTCDIPMCWTKD-GKTLRGARQOLSHEGHRAQLLI 3154
Db 2945 L-SKEVINVEVQAGDPLTECPIEDTSGVHTWRSQFGKQDQMDRAQSSD--KSKLYI 3001
QY 3155 TGATLQDSGRKVC---BAGGACSSSIVRVHARVVRFOEALKDLEVLGGAATLRCVLSS 3210
Db 3002 MCATPEDADSFCIAVNDAGAEAVFQVTVNTPKIFGDSFSTTEIVADITLPIPCRTEG 3061
QY 3211 VAAP-VKWCY-GNNVLR-PGDKYSLRQEGAMLELIVRNLRPQDSGRYCSFGDQTTSA-- 3265
Db 3062 IPPPEISWFLDGKPILEMPGVY---KQGD-LSLRIDNIKNPQOEGRYTCVCAENKAGRAEQ 3117
QY 3266 -TLTVTALPAQFIGKLRNKEATEGATLRCLEL-SKTAPVWEKSGSETLRDGRY--C 3319
Db 3118 DTYVEISEPFRVWASEVVRVVEGRQTTIRCEVGNPEPVVW-----LKGEFYTSDL 3171

Db 3378 VKDGCDAWKBPEDGGAEISHYVIERQDAATGATGSGKDNFHVDDLTQGEHYKPR 3437
QY 3223 ---VLBPCKYSLRQEGAML---ELV---VRNLRPODSGRYSCS 3257
Db 3438 KVAVNRHGDSDPLEAREAAIAKDPFRADKPGTPEIVDMKDHADLKWTPPADGGAPIE 3497
QY 3258 FGDQTTSATLVTALPAQFICKLRNKGATEGATLRLCELSKTAPVWRKGSSETLRDGR 3317
Db 3498 -GYLVXRTPSGDVNPVTVG---AGE-LTATV---DGLKPGQT 3533
QY 3318 YCLRODGMCELOIRGLAMVDAEYSCVCGERTSASLTIRPMPAHFIGRLRH---QE 3372
Db 3534 Y-----QFRVKNALNAGE-----STPDPSRTMVAKPRHLAPKINDFMFAOR 3576
QY 3373 SIEGATAT--LRCELSKAAPVEM-RKGRESLRDGRSLRQDGAVALCELOICGLAVADAGE 3429
Db 3577 VKAGOTLNFVNVGEPAKPIEWFLNGSPSSGGNTHIDNNTDNTKLTYSARADSGK 3636
QY 3430 YSCVC---GBERTSATLTVKALPAKFTFEGURNBEAVEGATAMLWCELSKVAPVWRKGP 3485
Db 3637 YKIVATNBSGKDEHEVDVNILDIQAPBEPGLRHKDIITKESVVLKXDE---PLDDGGSP 3691
QY 3486 ---ENLRDGRVILRQETRCCELOICGLAMADAGEY---LCVCGERTSATLTI-RA 3535
Db 3692 ITNVVVEKQEDGGRWVPCGETSDTSLKVNKL--SEGHEYKFRVAVNRQGTSAPLTSDHA 3749
QY 3536 LPARFIEDVKQAEAREGATVLOCELSAAPVWRKGSSETL---RDG---DRYSL- 3584
Db 3750 IVA---KNPDEPDAP---TDVTPVDMDKHVDLEWKPAPPDNGGAPIDAYIVE 3796
QY 3585 ---RDG--TKC---ELQIRGLAMADTGEYS---CYCQER 3614
Db 3797 KDKFEDWVECARVDGKTTKATADNLTPGETYQFRVAVNKAAGPKPSDPTGNVAKPRR 3856
QY 3615 TSAMLTVRA---LPKFTFEGURNBEATEGATVRLCELSKMAPVEMWKHE--TLRDGRH 3670
Db 3857 MAPKLNLAGLLDLRIKAGTPKIDIAPEG-----EPAPVAKWKANDATIDTGARA 3906
QY 3671 SLRQDGAARCELOIRGLVAEDAGEVLCMC---GKERTSAMLTVRAMPKSFIEGLURNEA- 3725
Db 3907 DVTNPTSSAIIHFSAVRGDTGVYKIIIVENEHGXDTAQCNVTVLDVDFGTPGEPGLKIDEIH 3966
QY 3726 TEGDTATLWCELSKAAAPVWRKGHETLR-----DGRHSLRQDGRSCELOIRGLAVVDA 3779
Db 3967 KEG-----CTLNWKPFDD-NGGTDVLYHIVKXDTSGTWOEVGTFPDCTAKNKLVP 4019
QY 3780 GEYSCVCGERTSA-TLTVRALPARFIE-----DVK-N 3810
Db 4020 KEYAF-----RVKAVNLGSGSKLEAEPIIAKNQFVDPDPVKPEVTDMDKORIDIKWN 4074
QY 3811 QEARREGATAVLQ--CELSKAAPVWRKGSSETLRGDRYSL--RODGTRECELOIHLGSLVAD 3866
Db 4075 PTANNGGAPVTGYIVEKKEKSAIWTAGKT--PGTTFADNLKPGVEYEFVRVAVNAAG 4132
QY 3867 TGEYSCVCGERTSATLTVRAPQVFPREPLQSLQAEESTATLQCE-LSEPTATVVMWSK 3925
Db 4133 PSDPSDPTDPTTKA---RYLKPKILITASRKIKIKAGFTNLEVDFIGAPDPTATWTVG 4188
QY 3926 GLQIQANGREPRRL---QGTAEVLVLDLOREDTGBYCTCGSQATSAITLTVTAAPVRF 3981
Db 4189 ---DSGAALAPELLVDAKSTTSIFFPSAKRADSGNY-----KLKV 4226
QY 3982 LRELQHQB-----VDEGGTAHLCELSRAGASVEMWKSGSLQLPFCAYQWV-----QD 4029
Db 4227 KNELGEDEAIFEVIVQDRPSA-----PEGPLEVSDVTKDSVLNWKPPKD 4271
QY 4030 GAABELLVRGVEQEDAGDYCTDGTHTQSMASLSVVRPFRPKTRLQSLBQETGDIALCC 4089
Db 4272 DGGAEISNVVVEKBDTKNTW-----VPSAFVT----- 4300
QY 4090 QLSDAESGAVVQWLKEGVELHAGPKYEMRSQGAATRELLIHOLEAKDTEGACVTTGOKTA 4149

Db 4301 ----GTSITVPKLTGEGHE-----YEFR--VVAENTFGSRSDSLNTDEPVLAKDPFGTP 4346
QY 4150 ASLRVTEPEVTVIRGLVDAEVTADSDVEFS-----CEVSAGATGQVQWC-L 4194
Db 4347 G--KGRPEI-----VD---TDNDHIDIKWDPPRDNGSGSPVDHYDIERKDAKTRWIKV 4395
QY 4195 QGLPIQSNNEVTEVAVRDGRIHTLRUKGVTPEDAGTVSFHLGNHASSAOLTVRAPEVITLE 4254
Db 4396 NTSPVCGTAFSTRVQKG--HTYEYRVAVNKAAGPGQPSDSSAAATAKAPMEAPKFDL 4453
QY 4255 PLQDVQLSEGQDASQCRLSRASGOEABWAL--GGVPLQANEMNDITVEQGLHLLTLHKV 4313
Db 4454 DGKEFRVAKAGEPLVITITFTASQPQDISWTEGGKPLAGVETTDQTK-----LVIPST 4507
QY 4314 TLEDAGTVSFHVTG--CSSEAOIKVTAKNTVVRGLEVENAELEGGBALFECOLSQPEVAH 4371
Db 4508 RESDSPVKIKAVNPFYGAENIKITV-----IDKGAPE-----NITYPAVSRH 4552
QY 4372 TWLLDDEPVRITSENAEVVFFENGLRHLLLLKNLRPDQDSCTVTFLLAGDMVTSATLTVGR 4431
Db 4553 TCTLNWDAPKODGGAEIAGYKIEYOEV-----GSQIWDKVPGLISGTAYTVRG-- 4600
QY 4432 LEILEPLKNAAVRAGAQAQARFTCTLSEAVPVGEASWYINGAAV--QPDDSDWTVTADGSH 4488
Db 4601 -----LEHCQOXYFRIRAEANAVGLSD---YCOGVFVWIKDPDP----- 4636
QY 4489 QALLRSQAHPHAGEVTFACRDVAVASARLTVLGLPDPPEDAEVVAHSSHTVTLSWAAPMS 4548
Db 4637 -----PGAPSTPEITGYDTNQVSLAWNPPRD 4662
QY 4549 DGGGLGCGYRVE--VKEGATGQWR--LCHELVPGBCVVDGLAPGETYFRVAAVGPVCGAG 4605
Db 4663 DGGSPILGVVBERPEKGG--GDWAPVPMVWKGTECIVPGLHEHETVQFRVAVNAAGH 4721
QY 4606 EPVHLPTVRLAEPKPPVPPQPSAPESPQVAAGEDVSLELEVAVABAGVIVHKGWRIQ 4665
Db 4722 EPSNGSEPTVC---RPVYKPGAPDAPRG-----KITKNSABLTWNR---PLRD 4765
QY 4666 GG-----REFVVSQGRQOQMLVIKGTAEQDQGEVHCG 4696
Db 4766 GAPIDGYIVEKKLGDNDWTRCNCKPVRDTAFEVKNLGEKE-----EVEFR 4812
QY 4697 LAQGSICPAATFQVALSPASVDEAPQPSLPPEAAQEGDLHLLWEALARKERMSREPTLD 4756
Db 4813 V-----JAVNSAGEGEPKPS-----DLVLEBQPCR-----PIFD 4843
QY 4757 --SISELPEEDGRS--QRLPQEAEEVAP--DLSEGYSTADELART-----GDADLSHTS 4804
Db 4844 INMLKIDTVRAGETQIRIPYAGGNPKPIIDLFNGNSPIFENERTVVDNPFVITTTG 4903
QY 4805 SDESRAGTPSLVITYLKAGR-----PG--TSPL-ASKVGAAPAP-SYKPOQ 4848
Db 4904 S-KRSDAG-PYKISATNKYGDCTKLVNVLDAPEGKTPGIRATDIQADAMTSLWRPCKD 4961
QY 4849 Q-----EPLAAVRPPLGDLSTKDLGDFSDMKAAVKIOAAPKGVKVRKEMQOEGPMESH 4902
Db 4962 NGGDATVYVWEKRTFGGDWVT--VGHEPVGTTLVRNLDANTPXYFRVRAENQYV----- 5014
QY 4903 TFGDTEAOVGDALRL-----ECVVASKADVRAARWLK---DG----- 4935
Db 5015 -----VGEFLETTDAIVAKNPPDTPGAPQGPVAVETSEAITLQWTRPDSGGAPIQ 5066
QY 4936 ---VE-----LTDGRHHHIDOLGDTCSLLIAGLDRADAGCYTCQVSNK 4976
Db 5067 GVVIKREYGVSTEWTKAAGFNILDTKRVGTGLTPKTYEFVRVAAATNAAGQGEYS----- 5120
QY 4977 FQGVTHSACVVVSGSESESSGCELDLDAFRRAARRLHRLFR--TKSPAESVDEELFSLA 5035
Db 5121 ----VNSVPTADNAPTFRKINMGMLTRDILAYAGERAKILVPPAASPA----PKVTFSK 5172
QY 5036 DEGPAPPEADWQTYREDEHFCIRFBALEARQAVTRFQEMFATIGIVEIKLVEQ-- 5093
Db 5173 GENKISPTDPRVKVEYSDFLATLTIEKSELTDGGLYFVELENSQSGDSASIRLKVWDKPA 5232

A:Accession: S07571
 A:Molecule type: DNA
 A:Residues: 792-6839 <BEN2>
 A:Cross-references: EMBL:X15423; NID:g6897; PIDN:CAA33463.1; PID:g6898
 A:Experimental source: var. Bristol
 R:Benian, G.M.; Kiff, J.B.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.
 Nature 342, 45-50, 1989
 A:Title: Sequence of an unusually large protein implicated in regulation of myosin activity
 A:Reference number: S06797; MUID:90044042; PMID:2812002
 A:Accession: S06797
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 806-1175,1178-1998,'Y',2000-3040,'I',3042-3335,'I',3337-5693,5696-6359,'I',6361-6839
 A:Cross-references: EMBL:X15423
 A:Experimental source: var. Bristol
 R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
 Genetics 134, 1097-1104, 1993
 A:Title: Additional sequence complexity in the muscle gene, unc-22, and its encoded protein
 A:Reference number: S57218; MUID:93387664; PMID:8397135
 A:Accession: S57218
 A:Molecule type: DNA
 A:Residues: 2-99,108-194,'Q',196-206;374-468;658-753 <BEN4>
 A:Cross-references: EMBL:X15423
 A:Experimental source: var. Bristol
 R:White, S.
 submitted to the EMBL Data Library, May 1996
 A:Reference number: Z20442
 A:Accession: Z27934
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 'MGIPGKCKQ',19-6839 <W12>
 A:Cross-references: EMBL:Z73897; PIDN:CAA98064.1; GSPDB:GN00022; CESP:ZK617.1a
 A:Experimental source: clone ZK617
 R:Harris, B.
 submitted to the EMBL Data Library, May 1996
 A:Reference number: Z20458
 A:Accession: Z28030
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 'MGIPGKCKQ',19-6839 <W12>
 A:Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CESP:ZK617.1a
 A:Experimental source: clone ZK829
 C:Comment: Lack of unc-22 leads to a constant twitching of the body muscles.
 C:Genetics:
 A:Gene: unc-22; CESP:ZK617.1a
 A:Map position: 4
 A:Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3; 601/3; 629/3; 669/3; 6776/1; 6808/3
 C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;
 C:Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine/threonine kinase
 F:806-898,899-990,991-1083,1084-1175,1178-1273,1474-1567,1770-1864,2066-2158,2358-2450,2451-2452,2453-2454,2455-2456,2457-2458,2459-2460,2461-2462,2463-2464,2465-2466,2467-2468,2469-2470,2471-2472,2473-2474,2475-2476,2477-2478,2479-2480,2481-2482,2483-2484,2485-2486,2487-2488,2489-2490,2491-2492,2493-2494,2495-2496,2497-2498,2499-2500,2501-2502,2503-2504,2505-2506,2507-2508,2509-2510,2511-2512,2513-2514,2515-2516,2517-2518,2519-2520,2521-2522,2523-2524,2525-2526,2527-2528,2529-2530,2531-2532,2533-2534,2535-2536,2537-2538,2539-2540,2541-2542,2543-2544,2545-2546,2547-2548,2549-2550,2551-2552,2553-2554,2555-2556,2557-2558,2559-2560,2561-2562,2563-2564,2565-2566,2567-2568,2569-2570,2571-2572,2573-2574,2575-2576,2577-2578,2579-2580,2581-2582,2583-2584,2585-2586,2587-2588,2589-2590,2591-2592,2593-2594,2595-2596,2597-2598,2599-2600,2601-2602,2603-2604,2605-2606,2607-2608,2609-2610,2611-2612,2613-2614,2615-2616,2617-2618,2619-2620,2621-2622,2623-2624,2625-2626,2627-2628,2629-2630,2631-2632,2633-2634,2635-2636,2637-2638,2639-2640,2641-2642,2643-2644,2645-2646,2647-2648,2649-2650,2651-2652,2653-2654,2655-2656,2657-2658,2659-2660,2661-2662,2663-2664,2665-2666,2667-2668,2669-2670,2671-2672,2673-2674,2675-2676,2677-2678,2679-2680,2681-2682,2683-2684,2685-2686,2687-2688,2689-2690,2691-2692,2693-2694,2695-2696,2697-2698,2699-2700,2701-2702,2703-2704,2705-2706,2707-2708,2709-2710,2711-2712,2713-2714,2715-2716,2717-2718,2719-2720,2721-2722,2723-2724,2725-2726,2727-2728,2729-2730,2731-2732,2733-2734,2735-2736,2737-2738,2739-2740,2741-2742,2743-2744,2745-2746,2747-2748,2749-2750,2751-2752,2753-2754,2755-2756,2757-2758,2759-2760,2761-2762,2763-2764,2765-2766,2767-2768,2769-2770,2771-2772,2773-2774,2775-2776,2777-2778,2779-2780,2781-2782,2783-2784,2785-2786,2787-2788,2789-2790,2791-2792,2793-2794,2795-2796,2797-2798,2799-2800,2801-2802,2803-2804,2805-2806,2807-2808,2809-2810,2811-2812,2813-2814,2815-2816,2817-2818,2819-2820,2821-2822,2823-2824,2825-2826,2827-2828,2829-2830,2831-2832,2833-2834,2835-2836,2837-2838,2839-2840,2841-2842,2843-2844,2845-2846,2847-2848,2849-2850,2851-2852,2853-2854,2855-2856,2857-2858,2859-2860,2861-2862,2863-2864,2865-2866,2867-2868,2869-2870,2871-2872,2873-2874,2875-2876,2877-2878,2879-2880,2881-2882,2883-2884,2885-2886,2887-2888,2889-2890,2891-2892,2893-2894,2895-2896,2897-2898,2899-2900,2901-2902,2903-2904,2905-2906,2907-2908,2909-2910,2911-2912,2913-2914,2915-2916,2917-2918,2919-2920,2921-2922,2923-2924,2925-2926,2927-2928,2929-2930,2931-2932,2933-2934,2935-2936,2937-2938,2939-2940,2941-2942,2943-2944,2945-2946,2947-2948,2949-2950,2951-2952,2953-2954,2955-2956,2957-2958,2959-2960,2961-2962,2963-2964,2965-2966,2967-2968,2969-2970,2971-2972,2973-2974,2975-2976,2977-2978,2979-2980,2981-2982,2983-2984,2985-2986,2987-2988,2989-2990,2991-2992,2993-2994,2995-2996,2997-2998,2999-3000,3001-3002,3003-3004,3005-3006,3007-3008,3009-3010,3011-3012,3013-3014,3015-3016,3017-3018,3019-3020,3021-3022,3023-3024,3025-3026,3027-3028,3029-3030,3031-3032,3033-3034,3035-3036,3037-3038,3039-3040,3041-3042,3043-3044,3045-3046,3047-3048,3049-3050,3051-3052,3053-3054,3055-3056,3057-3058,3059-3060,3061-3062,3063-3064,3065-3066,3067-3068,3069-3070,3071-3072,3073-3074,3075-3076,3077-3078,3079-3080,3081-3082,3083-3084,3085-3086,3087-3088,3089-3090,3091-3092,3093-3094,3095-3096,3097-3098,3099-3100,3101-3102,3103-3104,3105-3106,3107-3108,3109-3110,3111-3112,3113-3114,3115-3116,3117-3118,3119-3120,3121-3122,3123-3124,3125-3126,3127-3128,3129-3130,3131-3132,3133-3134,3135-3136,3137-3138,3139-3140,3141-3142,3143-3144,3145-3146,3147-3148,3149-3150,3151-3152,3153-3154,3155-3156,3157-3158,3159-3160,3161-3162,3163-3164,3165-3166,3167-3168,3169-3170,3171-3172,3173-3174,3175-3176,3177-3178,3179-3180,3181-3182,3183-3184,3185-3186,3187-3188,3189-3190,3191-3192,3193-3194,3195-3196,3197-3198,3199-3200,3201-3202,3203-3204,3205-3206,3207-3208,3209-3210,3211-3212,3213-3214,3215-3216,3217-3218,3219-3220,3221-3222,3223-3224,3225-3226,3227-3228,3229-3230,3231-3232,3233-3234,3235-3236,3237-3238,3239-3240,3241-3242,3243-3244,3245-3246,3247-3248,3249-3250,3251-3252,3253-3254,3255-3256,3257-3258,3259-3260,3261-3262,3263-3264,3265-3266,3267-3268,3269-3270,3271-3272,3273-3274,3275-3276,3277-3278,3279-3280,3281-3282,3283-3284,3285-3286,3287-3288,3289-3290,3291-3292,3293-3294,3295-3296,3297-3298,3299-3300,3301-3302,3303-3304,3305-3306,3307-3308,3309-3310,3311-3312,3313-3314,3315-3316,3317-3318,3319-3320,3321-3322,3323-3324,3325-3326,3327-3328,3329-3330,3331-3332,3333-3334,3335-3336,3337-3338,3339-3340,3341-3342,3343-3344,3345-3346,3347-3348,3349-3350,3351-3352,3353-3354,3355-3356,3357-3358,3359-3360,3361-3362,3363-3364,3365-3366,3367-3368,3369-3370,3371-3372,3373-3374,3375-3376,3377-3378,3379-3380,3381-3382,3383-3384,3385-3386,3387-3388,3389-3390,3391-3392,3393-3394,3395-3396,3397-3398,3399-3400,3401-3402,3403-3404,3405-3406,3407-3408,3409-3410,3411-3412,3413-3414,3415-3416,3417-3418,3419-3420,3421-3422,3423-3424,3425-3426,3427-3428,3429-3430,3431-3432,3433-3434,3435-3436,3437-3438,3439-3440,3441-3442,3443-3444,3445-3446,3447-3448,3449-3450,3451-3452,3453-3454,3455-3456,3457-3458,3459-3460,3461-3462,3463-3464,3465-3466,3467-3468,3469-3470,3471-3472,3473-3474,3475-3476,3477-3478,3479-3480,3481-3482,3483-3484,3485-3486,3487-3488,3489-3490,3491-3492,3493-3494,3495-3496,3497-3498,3499-3500,3501-3502,3503-3504,3505-3506,3507-3508,3509-3510,3511-3512,3513-3514,3515-3516,3517-3518,3519-3520,3521-3522,3523-3524,3525-3526,3527-3528,3529-3530,3531-3532,3533-3534,3535-3536,3537-3538,3539-3540,3541-3542,3543-3544,3545-3546,3547-3548,3549-3550,3551-3552,3553-3554,3555-3556,3557-3558,3559-3560,3561-3562,3563-3564,3565-3566,3567-3568,3569-3570,3571-3572,3573-3574,3575-3576,3577-3578,3579-3580,3581-3582,3583-3584,3585-3586,3587-3588,3589-3590,3591-3592,3593-3594,3595-3596,3597-3598,3599-3600,3601-3602,3603-3604,3605-3606,3607-3608,3609-3610,3611-3612,3613-3614,3615-3616,3617-3618,3619-3620,3621-3622,3623-3624,3625-3626,3627-3628,3629-3630,3631-3632,3633-3634,3635-3636,3637-3638,3639-3640,3641-3642,3643-3644,3645-3646,3647-3648,3649-3650,3651-3652,3653-3654,3655-3656,3657-3658,3659-3660,3661-3662,3663-3664,3665-3666,3667-3668,3669-3670,3671-3672,3673-3674,3675-3676,3677-3678,3679-3680,3681-3682,3683-3684,3685-3686,3687-3688,3689-3690,3691-3692,3693-3694,3695-3696,3697-3698,3699-3700,3701-3702,3703-3704,3705-3706,3707-3708,3709-3710,3711-3712,3713-3714,3715-3716,3717-3718,3719-3720,3721-3722,3723-3724,3725-3726,3727-3728,3729-3730,3731-3732,3733-3734,3735-3736,3737-3738,3739-3740,3741-3742,3743-3744,3745-3746,3747-3748,3749-3750,3751-3752,3753-3754,3755-3756,3757-3758,3759-3760,3761-3762,3763-3764,3765-3766,3767-3768,3769-3770,3771-3772,3773-3774,3775-3776,3777-3778,3779-3780,3781-3782,3783-3784,3785-3786,3787-3788,3789-3790,3791-3792,3793-3794,3795-3796,3797-3798,3799-3800,3801-3802,3803-3804,3805-3806,3807-3808,3809-3810,3811-3812,3813-3814,3815-3816,3817-3818,3819-3820,3821-3822,3823-3824,3825-3826,3827-3828,3829-3830,3831-3832,3833-3834,3835-3836,3837-3838,3839-3840,3841-3842,3843-3844,3845-3846,3847-3848,3849-3850,3851-3852,3853-3854,3855-3856,3857-3858,3859-3860,3861-3862,3863-3864,3865-3866,3867-3868,3869-3870,3871-3872,3873-3874,3875-3876,3877-3878,3879-3880,3881-3882,3883-3884,3885-3886,3887-3888,3889-3890,3891-3892,3893-3894,3895-3896,3897-3898,3899-3900,3901-3902,3903-3904,3905-3906,3907-3908,3909-3910,3911-3912,3913-3914,3915-3916,3917-3918,3919-3920,3921-3922,3923-3924,3925-3926,3927-3928,3929-3930,3931-3932,3933-3934,3935-3936,3937-3938,3939-3940,3941-3942,3943-3944,3945-3946,3947-3948,3949-3950,3951-3952,3953-3954,3955-3956,3957-3958,3959-3960,3961-3962,3963-3964,3965-3966,3967-3968,3969-3970,3971-3972,3973-3974,3975-3976,3977-3978,3979-3980,3981-3982,3983-3984,3985-3986,3987-3988,3989-3990,3991-3992,3993-3994,3995-3996,3997-3998,3999-4000,4001-4002,4003-4004,4005-4006,4007-4008,4009-4010,4011-4012,4013-4014,4015-4016,4017-4018,4019-4020,4021-4022,4023-4024,4025-4026,4027-4028,4029-4030,4031-4032,4033-4034,4035-4036,4037-4038,4039-4040,4041-4042,4043-4044,4045-4046,4047-4048,4049-4050,4051-4052,4053-4054,4055-4056,4057-4058,4059-4060,4061-4062,4063-4064,4065-4066,4067-4068,4069-4070,4071-4072,4073-4074,4075-4076,4077-4078,4079-4080,4081-4082,4083-4084,4085-4086,4087-4088,4089-4090,4091-4092,4093-4094,4095-4096,4097-4098,4099-4100,4101-4102,4103-4104,4105-4106,4107-4108,4109-4110,4111-4112,4113-4114,4115-4116,4117-4118,4119-4120,4121-4122,4123-4124,4125-4126,4127-4128,4129-4130,4131-4132,4133-4134,4135-4136,4137-4138,4139-4140,4141-4142,4143-4144,4145-4146,4147-4148,4149-4150,4151-4152,4153-4154,4155-4156,4157-4158,4159-4160,4161-4162,4163-4164,4165-4166,4167-4168,4169-4170,4171-4172,4173-4174,4175-4176,4177-4178,4179-4180,4181-4182,4183-4184,4185-4186,4187-4188,4189-4190,4191-4192,4193-4194,4195-4196,4197-4198,4199-4200,4201-4202,4203-4204,4205-4206,4207-4208,4209-4210,4211-4212,4213-4214,4215-4216,4217-4218,4219-4220,4221-4222,4223-4224,4225-4226,4227-4228,4229-4230,4231-4232,4233-4234,4235-4236,4237-4238,4239-4240,4241-4242,4243-4244,4245-4246,4247-4248,4249-4250,4251-4252,4253-4254,4255-4256,4257-4258,4259-4260,4261-4262,4263-4264,4265-4266,4267-4268,4269-4270,4271-4272,4273-4274,4275-4276,4277-4278,4279-4280,4281-4282,4283-4284,4285-4286,4287-4288,4289-4290,4291-4292,4293-4294,4295-4296,4297-4298,4299-4300,4301-4302,4303-4304,4305-4306,4307-4308,4309-4310,4311-4312,4313-4314,4315-4316,4317-4318,4319-4320,4321-4322,4323-4324,4325-4326,4327-4328,4329-4330,4331-4332,4333-4334,4335-4336,4337-4338,4339-4340,4341-4342,4343-4344,4345-4346,4347-4348,4349-4350,4351-4352,4353-4354,4355-4356,4357-4358,4359-4360,4361-4362,4363-4364,4365-4366,4367-4368,4369-4370,4371-4372,4373-4374,4375-4376,4377-4378,4379-4380,4381-4382,4383-4384,4385-4386,4387-4388,4389-4390,4391-4392,4393-4394,4395-4396,4397-4398,4399-4400,4401-4402,4403-4404,4405-4406,4407-4408,4409-4410,4411-4412,4413-4414,4415-4416,4417-4418,4419-4420,4421-4422,4423-4424,4425-4426,4427-4428,4429-4430,4431-4432,4433-4434,4435-4436,4437-4438,4439-4440,4441-4442,4443-4444,4445-4446,4447-4448,4449-4450,4451-4452,4453-4454,4455-4456,4457-4458,4459-4460,4461-4462,4463-4464,4465-4466,4467-4468,4469-4470,4471-4472,4473-4474,4475-4476,4477-4478,4479-4480,4481-4482,4483-4484,4485-4486,4487-4488,4489-4490,4491-4492,4493-4494,4495-4496,4497-4498,4499-4500,4501-4502,4503-4504,4505-4506,4507-4508,4509-4510,4511-4512,4513-4514,4515-4516,4517-4518,4519-4520,4521-4522,4523-4524,4525-4526,4527-4528,4529-4530,4531-4532,4533-4534,4535-4536,4537-4538,4539-4540,4541-4542,4543-4544,4545-4546,4547-4548,4549-4550,4551-4552,4553-4554,4555-4556,4557-4558,4559-4560,4561-4562,4563-4564,4565-4566,4567-4568,4569-4570,4571-4572,4573-4574,4575-4576,4577-4578,4579-4580,4581-4582,4583-4584,4585-4586,4587-4588,4589-4590,4591-4592,4593-4594,4595-4596,4597-4598,4599-4600,4601-4602,4603-4604,4605-4606,4607-4608,4609-4610,4611-4612,4613-4614,4615-4616,4617-4618,4619-4620,4621-4622,4623-4624,4625-4626,4627-4628,4629-4630

Db 786 ---GGEKPM-TEAERQSFPKCKVEKWDIPPEKTVQOQVDKICEWKCTYSRPNAKI 840
Qy 1382 TWYKDGKLSSS---KYRMEAVGCTRLRVQOQACQADTGEYSCEAGQRLSFLSDVAP 1438
Db 841 RWYKORKEIFSGGLKYKIVIEKNVCT--LIINNEVDVDTGKYTCANGVPHQAULTVLEP 898
Qy 1439 KVFPAKEQVUREVQOAGASTTILSCEVAQOATEVMYKQKLSFSK---VRMEAVG- 1494
Db 899 PMKYSFLNPLENTQEIYRTKQAVLTCKVNTFRAPLVYRGSKAIOEGDPRFIIEKDAVR 958
Qy 1495 CTRRLVQOQACADAGEYSCEAGSQRLS-FHLHVAEPKAVFAKEQBPASREVOAAGTSAT 1553
Db 959 CT--LTIKEVEDDOAEWTARITQDVFSKVQVYVEEPRHTFVPMKSKVNESDL---AT 1013
Qy 1554 LSCEVAQOATEVMYKQKLS--SSSKVRMEAVGCTRLRVQOQACADAGEYSKAGDQR 1612
Db 1014 LETDVNDKDAEVMWHDGRIIDIVGVFKVBESSRKRLIINGARIEDHGEYKCTTKDDR 1073
Qy 1613 LSFHLHVAEPKVPVFAKEQ--PAHREVOAEAGASATLSCEVAQAOTEVMTYKDGKLS-- 1668
Db 1074 TMAQI-----IVAKNKFIVALKDTEVIEKDDVTLKQYKDTKTPGIWFRANGQISSMP 1127
Qy 1669 SSKVRMEAVGCTRLRVQOQACADAGEYSCEAGQRLSFLHVAELE--FOISERPCK-- 1724
Db 1128 GGFETQSRNGTHTLKIGKIEWNEADVYEDQAGLRGSCNVTVLEAKRPILNWKPKKIE 1187
Qy 1725 ---REPLVVKHEHDIILTATLATPSAATVWLKQGVETRRSKRH--ETASQGDTHLTIVH 1779
Db 1188 AKAGEPCVVVFFQIKGTRR-GDPKAQI---LKNKGPIDEMRKLEVEIHKDDVAEIVFK 1243
Qy 1780 GAQVLDSAIYSCRVG-----AEGQDF-----PV 1802
Db 1244 NPQLADTKWALELGNAGTALAPFELFVKDKPKPKPLKNTVTAEGDLVWGTDPDP 1303
Qy 1803 QVEVAAKFCRLLE---PVCGLGTVTLACELSPACABVVR-----CGNTOPRV 1850
Db 1304 EGAPYKAYIIEQEGRSNGNAKVGETKGTDFKVKDLKEH-GEYKFRVYKALNECGLSDPLT 1362
Qy 1851 GKRFORMAEGPVRSLTVGLRAEADAGEVVCSSRDHTSAQLTVSVPRVFMKSLSTWA 1910
Db 1363 GE--SVLAKNP-----YGVFGPKMMDAIDVDKDHCT-----LAMEPP 1398
Qy 1911 EGGGATFCQVSPSDVAVWFRDGLALQPSK-----FATSQSGASHSTISDLVEDA 1965
Db 1399 BEDGGAPI-----TGYIIEERKEKSEKWHQVGTQPCCELTDRKVVEDK 1443
Qy 1966 GOI--TVEAGSSSAALAVREAPVLFK--KLEPQ-----TVEERSSTVLELT-R 2013
Db 1444 EYLRYVKAVNKAAGPDCDHGKPIKOKAKKASPEFTGGGIKDLRKVGETIKYDVPISGE 1503
Qy 2014 PMPELRWTRNATALAPGKNVETIHAEGARHRLVHNVGFPADRGFF-----GCETPDD 2064
Db 1504 PLPECLWVNGKPLKAVGRVMSSESGKHIMKIENAVRADSGKFTITLKNSSGSCD--- 1559
Qy 2065 KTOAKLTVMROVRLVGLQAVEAREQGTATVE-----VOLSHADVDGS 2108
Db 1560 -STATVTVGRTPEPKPLDADVCADG-ATLSMWNPPDDGDPFLTGIVBEAQMDNKGK 1617
Qy 2109 WTRDGLRFOQGTCHLVRGPMHTLTLSGLRPEDSGLMVFAEGVHTSARLVWVTELPVSV 2168
Db 1618 YIEVG-----KVDPNITTLKVNGLR--NKGNYKFRVAVNNEGE----- 1654
Qy 2169 SRPLO-DVVYTEKEKVTLECELSRP-----NVDVRW-----LKDQ-----VELRAG 2208
Db 1655 SEPLSADQYQIKDPWDEBPGKPRPEITDFADRIDIAWEPHKGAGAPIBEYIVVEVRDP 1714
Qy 2209 KTMATAAQACASLTIYCEPADQGVYVCDADAGSASVVKVQRTYTLIVRRVLAEDAG 2268
Db 1715 DTK-----EMKE-----VKRPDNTNASTISGLKEGKEVQFRVAVNKAAGP 1754
Qy 2269 EIQFVAENAESRAQL-RVKELFVTLVRPLRDKIAMEKRGVLECOVSRASA----- 2318

Db 1755 Q-----PSEPSKQJLAKPKFIPAWL-----KHDNLKSITVKAGATVRWEVKIGG 1798
Qy 2319 ---QVRWPKGQELQPGPKYELVSD-----GLYRKLIISDVHAED 2355
Db 1799 EPIPEVKFKGNQOLENG-IQLTIDTRKNEHTILCIPSAMSDVGEYR-LTVKNSHGAD 1855
Qy 2356 E-----DTYTCADAGVKTSAQFVVEOSITIVRGLQDVTVMBP-----AP 2395
Db 1856 EKANUTVLDRFSKNGPLEVSDVF---EDNLNLSWKPPDDDDGGGPIEYVEVEKLDATG 1912
Qy 2396 AMFEC-----RPPKWLIGKTVLQ-----AGNVGLE-----Q 2429
Db 1913 RWVPCAKVKDTHKAHIDGLKKGQTYQFRVAVNKEGASDALSTDKDTKAKNPYDEPKTGT 1972
Qy 2405 PSV-----RPPKWLIGKTVLQ-----AGNVGLE-----Q 2429
Db 1973 PDVVDHADRVSLEWEPKSGDGAPIQTQVIEKKKGHDWQEOCKVSGDQNAEILGUK 2032
Qy 2430 EGTVHMLRLRSTCTMTGPHVHTVGKSSARLVVSDIPVLTTRPLEKPTKRELQSVL- 2488
Db 2033 EGEEYQFRYKAV--NKAGP-----GEASDPSKVVAK-----PRNLKPMIDREAMKTI 2080
Qy 2489 ---SCDF-----RPAPKAVOMYKDDTPLSPSEKFKMSLEGOMAELELRLIMPADAGVY 2538
Db 2081 KVGNDVFEFVVRGEPKPK-EMIFNEKEVD-DQIRIESEDYKTRFVLRGATRKHAGLY 2138
Qy 2539 RCOA-----GSAHSSTEVTEAREVTVTQLODAEATEGWSFSCELSHEDSEVMSL-- 2592
Db 2139 TLTATWASSDXHSVEVILGKPSPLGLEVSNVY------DRADLEWKVPE 2186
Qy 2593 ---NGMPLYNDSFHEISHK-----GRRHTLVLSIORADAGIVRASSLKVSTARS 2640
Db 2187 DGGGAPI---DHYEIEKMDLATRWVPCGRSET-----TKTTPVNLQPGHEY 2230
Qy 2641 EVRVKPVVFLKALDLDLSAE---ESGTLALQCEVSDPEAHVVRKDGVCOLG----- 2687
Db 2231 KFRVAVNKEGSDPLTTNTAILAKNPYEVPKVKDPKL-VVNDKDHVDLWKNAPDDGA 2289
Qy 2688 ---PSD-----KYDFLHTAGTGLVHVHDVSPED 2712
Db 2290 PTEAFVIEKKNGRWEALVPGDQKTATVPNLKEGEBYQFRISARKAGTGDPSDPSD 2349
Qy 2713 AGLYTCHVSEETRAVRVHDHLVHGITKLTMEVLEGESCFEVLSHESADPAMTV 2772
Db 2350 ---RVVAKPRNLAPRIH-----REDLSDTTKVKGATLKFIVHIDGEPADP-VTWBF 2396
Qy 2773 GGTIVGSS-----SRF-----OATQRGKYLIVVREAAAPSADAGEVVFVSRGLTSK- 2817
Db 2397 NGKIGESKAQIENEPYISRFPALPKALRKQSKYITITAININGTDSVTINIKVSKPTKP 2456
Qy 2818 ---ASLIVRERPAAIKPLEDQ-----WV----- 2838
Db 2457 KGPIETDVFEDRATLCKWPPEDDGGEPFEFYIEKQNTKDGIVWPCGRSGDTHFTVDSL 2516
Qy 2839 ---APGEDVELRCEL-----SRAGTP-----VHV--- 2859
Db 2517 NKGDHYKFRVAVNSGSDPLETETDILAKPPRDPGRPEPTDMDSDHVDLKNWDP 2576
Qy 2860 LKDRKA-----IRKSQYD-----VVCETGMAMLVIRGASIKXDAGEYTCV- 2900
Db 2577 LSDGGAPIBEYQIEKRTKYGRWEPAITVPGQTTATV--PDLTNPBEEYFRVAVNKGGP 2634
Qy 2901 ---EASKSTASHVEEKANCFTTELTNLOVEKGTAVFTCKTE-HPAATVTVWR-KGLLE 2954
Db 2635 SDPSDASKAVIAKPRNLKPHIDRALKNLTIKAGQISIFDVVSGSPAPTVTWHWPDNRE 2694
Qy 2955 LRASGHQPSQEGTLRLTISALEKADSDTYTCIDIGQAQSRRAQLLVQGRVHIE----- 3009
Db 2695 IRNGRKVLDPNPEYQSKLVVKQMERGDSGFTTIKAVNANGEDBATV---KINVIDKPTSP 2751
Qy 3010 ---DLEDVD-----VQEGSSAT----- 3023
Db 2752 NGPLDVSVRHDVTLNWRAPDDDDGIPPIENYVIEKYDTASGRWVPAKVAGDKTTAVVD 2811

| | | | |
|----|------|--|------|
| Qy | 3024 | -----FRCKISPNATY-----PVHWF-----LDKTPPL-----HANELNEID----- | 3054 |
| Db | 2812 | GLIPGHEYKFRVAANAEGESDPLETFGTTLAKDPFKPGKTNAPETIDWDKOHVDLEWK | 2871 |
| Qy | 3055 | --AQPGG-----YHVLTILRQLA-----LKDSCTIVF-----EAG | 3081 |
| Db | 2872 | PPANDGGAPIEYEVEMKDPSPFWNDVAHPAGQTNATVGNLKEGSKYBFRIRAKNGK | 2931 |
| Qy | 3082 | ----DQASAAALRVTKEPSFSR--ELTDAITTEGEDTLVCETSTCDIP--MCMTKDGKTL | 3135 |
| Db | 2932 | LGDPSDSASAVAKARNVPVIDRNSIQEIKVKAGQDFSLNIPVSGEPTIITWTFEGTPV | 2991 |
| Qy | 3136 | RGSNARCOLSHEGHPAQLIITGATLODSGRY--KCEAGACSSSIVRVHA-----RPV | 3185 |
| Db | 2992 | ESDDRMKLNNEDGKTKFHVXRALRSDTGTVIIKAENENGTDTAEVKVTVLDPHSPSPRGL | 3051 |
| Qy | 3186 | RFQEAALDKLEVL-----EGGAATLRCVLSSV--AAPVKV-----CYGN | 3221 |
| Db | 3052 | DVTINIVKGCGLAWKEPEDDGGAEISHYVIEKQDAATGRWTACGESKDTNFHVDLDTQGH | 3111 |
| Qy | 3222 | N-----VLRPGKYSLROEGAML-----ELV-----VRNLRPODSG | 3252 |
| Db | 3112 | EYKFRVKAVNRHGSDPLEAREAITAKDPFDRADKPGTPEIVDWDKHADLKWTPPADDG | 3171 |
| Qy | 3253 | RYSCSFGDQITSAITLVTPALPAQPIGKLNRKEATEGATATLRCELSKTAPVWRKGSETL | 3312 |
| Db | 3172 | GAPIE--GYLVEMRTPSGDWVFAVTVG-----AGE--LTAIV-----DGL | 3207 |
| Qy | 3313 | RDGDRYCLRQGMACELOIRGLAMVDAEYSVCVGEERTSASLTIREMPAHFICRLRH-- | 3370 |
| Db | 3208 | KPGQTY-----QFRVKALNKAGE-----STPSPRTWAKPRHLAPKINRDM | 3250 |
| Qy | 3371 | --QESIEGATAT--LRCELSKAAPVEM--RKGRESLRDGRHSRLRQDCAVCELOICGLAV | 3424 |
| Db | 3251 | FVAQRVKAGQTLNFDNVNVEGEPAPKIEFWLNGPLSGGNTHDNNNTNKLTKTSTAR | 3310 |
| Qy | 3425 | ADAGEYSCVC-----GEERTSATITVKALPAKFTFEGLRNEEAVEGATAMLCELSKVAPVE | 3480 |
| Db | 3311 | ADSGKYIVATNESGKDEHEVDVNILDIPGAPGEPLEHKDITKESVVLKWE-----PLD | 3365 |
| Qy | 3481 | WRKGP-----ENLRDGRVILRQEGTRCELOICGLAMADAGEY--LCVCGQERTSATL | 3531 |
| Db | 3366 | DGGSPIYTYVVEKQDGRVVPCCGSTDTSLKYNKL--SEGHYKFRVAVNRQGTSAPL | 3423 |
| Qy | 3532 | TI--RALPARFIEDVQNEAREGATAVLOCELNSAAPVWEVRKGSETL-----RDG--D | 3580 |
| Db | 3424 | TSDBAIVA-----KNPFDPEADP-----TDVTPVDWDKHVDLEWKPPANDGGAPID | 3470 |
| Qy | 3581 | RYSL-----RQDG--TKC-----ELQIRGLAMADTGEYS-----CV | 3609 |
| Db | 3471 | AYIVEKKDKFGDMVECARVDGKTKTATADNLTPGETVQFRVKAVNKGAPKPSDPTGNVV | 3530 |
| Qy | 3610 | CGQERTSAMLTVRA--LPIKFTFEGLRNEAREGATAVLRCELSKAAPVEMWKGHE--TLR | 3665 |
| Db | 3531 | AKPFRMAPKUNLAGLLDLRIKAGTPKIDTAFEQ-----EPAPVAKWANDATID | 3580 |
| Qy | 3666 | DGRHSRLRQDARGCELOIRGLVAEDAGEYLCMC--GKERTSAMLTVRAMPSKFTIEGLR | 3721 |
| Db | 3581 | TGARADVNTPTSSAIHIFSAVRGDTGVYKIIIVENEHKGDTAQCNVTLDVPCTPEGPLK | 3640 |
| Qy | 3722 | NEBA--TEGDTATLWCELSKAAPVWRKGSHTLRL-----DGRHSRLRQDGSRCOLOIRGL | 3774 |
| Db | 3641 | IDEIHKEG-----CTLNWKPPTD--NGGTDVLHYIIVEKMDTSRGTQWVGTFDPCDAKVN | 3693 |
| Qy | 3775 | AVVDAGEYSCVCGQERTSA--TLTVRALPARFIE----- | 3806 |
| Db | 3694 | KLVPGKEYAF-----RVKAVNLQESKPLEAEPIITAKNQDFVDPDPVDPKFTVMDWKDRI | 3748 |
| Qy | 3807 | DVK--NOEAREGATAVLO--CELSKAAPVWEVRKGSETLRGGDRVSL--RQDGRTRCELOIHG | 3861 |
| Db | 3749 | DIKWNPTANNNGGAPVTGYIIVEKKEKGSAMWTEAGKT--PGTTFESADNLKPGVIEYEPRLTA | 3806 |

| | | | |
|----|------|---|------|
| Qy | 3862 | LSVADTGEYCVCCQBERTSATLTVRAPOPVREBPLOSLOABEGSTATLOCE-LSEPTATV | 3920 |
| Db | 3807 | VNAAGSPDPDPDQPTTKA-----RYLKPKRILTASRKIKKAGFTINLEVDIFGADPDTA | 3862 |
| Qy | 3921 | VWSKGGIQLQANGRRERL-----QGCTAEVLVLQLOREDTGEVTCCTCGSQAATSATLTVTA | 3976 |
| Db | 3863 | TWTVG-----DSGAALAPELLVDAKSSTTSFFPSAKRADSGNY----- | 3901 |
| Qy | 3977 | APVRFJLRELQHE-----YDEGGTAHLCCELSRAGASVWRKGSLLQLPFCAKYQMV--- | 4027 |
| Db | 3902 | KLKVKNELGEDEAIPEVIVODRPSA-----PEGLEVSDVTWKDSCVLNW | 3945 |
| Qy | 4028 | ---ODGAAAEALLVRGVEQEDAGDYTCDTGHTQSMASLSVVRPPKPKFKTRLQSLQEQETGDI | 4084 |
| Db | 3946 | KPPKDDGGAIEISNYVKEKRTKNTW-----VPSAVFT----- | 3979 |
| Qy | 4085 | ARLCCQLSDAESGAAVQWLKEGVHELHAGPKYMRSQCATRELLIHOLEAKDTGEYACVTVG | 4144 |
| Db | 3980 | -----GTSITVPKLTGHE-----YEFR-----VMAENTFGSRSDSLINTDEPVLAKD | 4020 |
| Qy | 4145 | GOKTAASLRVTEBEVTVIRGLVAEVTADBDVEFS-----CEVSRAQATGV | 4190 |
| Db | 4021 | PFGTGP--KPEGRPEI-----VD---TDNDHIDIKWDPDRDNGSGPVHDYHIERDKDAKTG | 4069 |
| Qy | 4191 | QWC--LOGLPLQSNTEVEVARDRGRIHTLRLKGYTPEDAGTVSPHLGNHASSAQLTVRAPE | 4249 |
| Db | 4070 | RWKVNTSPQGTAFSDTRVQKG--HTYERYVVAVNKAGGQGDSSAAATAKPMHEAPK | 4127 |
| Qy | 4250 | VTILEPDQYQLSEGDQASFCQLSRASQOEARWAL--GGVPLOANEMNDITVBEQGTLLHL | 4308 |
| Db | 4128 | FDLDLCKGERVRKAGEPLVITIPFTASPODISWTKEGGKPLAGVETTDSQTK-----L | 4181 |
| Qy | 4309 | TLHKVLTLEDAGTVSFHVGTT--CSSEAQLKVTAKNTVVRGLENVLEALEGGEALFEQOLSOP | 4366 |
| Db | 4182 | VIPSTRRSDSGPVKIKAVNPYGEAEANIKITV-----IDKPGAPE-----NITVP | 4226 |
| Qy | 4367 | EVAAHTWLLDDBEVPTSENAAVVFFENGLRHLLLLKLNLRPOQSCRVTFLAGDMVTSAFLT | 4426 |
| Db | 4227 | AVSRHTCTLNWDAPKDDGGAIEAGYKIEYEV-----GSQIWDKVPGLISGTAYT | 4276 |
| Qy | 4427 | VRGWRLEILEPLKNAAVRAGAQAARFTCTLSEAVPVGESWYINGAAV---QPDDSDWTVT | 4483 |
| Db | 4277 | VRG-----LEHGQQYRFRIRAEANAVGLSD---YQQGVPVVWKDPFDP----- | 4315 |
| Qy | 4484 | ADGSHQALLRSQAOPHHAGEVTFACRDASARLTVLGLDPDPEDAEVVAHSSHTVTLISW | 4543 |
| Db | 4316 | -----PGAPSPETITGYDTNQVSLAW | 4336 |
| Qy | 4544 | AAPMSDGGGLCGYRYE--VKEGATQWR--LCHELVPGPCCVVVDGLAPGETYRFRVAAYG | 4600 |
| Db | 4337 | NPPRDDGGSPILGYVVERPEKRG--GDWAPVKMPVMVKTECIVPGLHENETYQFRVAVN | 4395 |
| Qy | 4601 | PVGAGEBVHLPOVRLAEPKPVPPOPSPAPESQVAAAGEDVSLLEVVAAEGEVTMHKGM | 4660 |
| Db | 4396 | AAGHGSPNSGESEVTC-----RPVVEKPGAPDAFRVG-----KITKNSAELTWNR--- | 4440 |
| Qy | 4661 | ERIQPG-----RFEVWSQGRQOQMLVIXGFTAEODG | 4691 |
| Db | 4441 | PLRDGGAPIDGYIVEKKLGDNDWTRCNDKPRVDTAPEVKNLGEK----- | 4486 |
| Qy | 4692 | EYHCGLAQSGICPAATTFQVALSPASVDPAQPSLPPEAAQSGDLHLLWEALARKMRSR | 4751 |
| Db | 4487 | EYEBRV-----IAVNSAGEGEPKSPS-----DLVLISEQPGR----- | 4518 |
| Qy | 4752 | EPTLD--SISELPEEDGRS--ORLPOAEAEVAP--DLSEGYSTABELART-----GDAD | 4799 |
| Db | 4519 | PIFDLNNLKDITVRAGETIQIRIPVAGGNPKPIIDLFGNSSPIFENERTVVDVNPGEIV | 4577 |
| Qy | 4800 | LSHTSDSDSERAQTPSLVYLKKAGR-----PG---TSPL-ASKVGAPAAP--SV | 4843 |
| Db | 4578 | ITTTGS--KRSdag--PYKISATNKYIGKDTCKLVNFVLDAPGKPTGIRATDIOADAMTLSW | 4635 |
| Qy | 4844 | KPOOOC-----EPLAAVPPPLGDLSTKDLGPPSDMKAAVKTOAAFPKGYKVRKEMKOEGB | 4897 |

```

Db 4636 RPPKNGDGAITNVVBEKTPGGDWVT--VGHVPVGTTLVRNLDANTPYEVRVRAENQYQ 4693
QY 4898 PMFSHTFGDTBAQVGDALRL-----ECWVASKADVRAERWLK---DG 4935
Db 4694 -----VGEPLETDDAIVAXNPFDTGAPQPEAVETSEBEAITLQWTRPTSDG 4740
QY 4936 -----VE-----LTDGRHHHIDOLGDTGCSLLIAGLDRADAGCYTC 4971
Db 4741 GAPIQGVYIEKREVGSTEWTKAAGFNILDTKHRVTGLTPKTYEYFRVAAYNAAGQGEYS- 4799
QY 4972 QVSNKFGVTHSACVWVSGSESESSGGELDDAFRAARLRLHFR-TKSPAESVDEE 5030
Db 4900 -----VNSVPITADNAPTRPKINNGMLTRDILAYAGERAKILVPPFAASA- 4846
QY 5031 LFLSADGGPABPEPADQWYREDEHFCIRFEALTEARQAVTRFQMFATLIGIVEIKL 5090
Db 4847 VTFSGKGNKISPTDRVKVEYSPLATLITEKSELTDGLYFVELENSQSDGSASIRLKV 4906
QY 5091 VEQ--GRRVEMCISKETAP-----VVPPEPLPSLLTSDAAPV--FLTELQNEVQDG 5140
Db 4907 VDKPASQHIRV--EDIAPDCCTLYWMPFS-----SGGSPITNYIVEKLDLRHSDG 4956
QY 5141 -----YVVSFDC-----VVTGQP- 5153
Db 4957 KWEKVSFVRNLNVTGGLIKDNRYRFRVRAETQYGVSEPCELADVVVAKYQFEVNPQE 5016
QY 5154 MPSV-----RWFQDKLLEDDHYMINEDQ 5178
Db 5017 APTVRDKDSTWAELEWPPRDRGSKIIIGYQVQVADTSSGWINAKMDLSQCHARTGLR 5076
QY 5179 QGGHOLIITAVPADMGVYRCLAENSMGVS--STKAELR-----VDLTS-- 5220
Db 5077 QNGE-----PEFRIIAKNAAGFKSPSPSERCQLKSRFGPPGPIHVGAKSIG 5124
QY 5221 -----TDYDTAADATSSSVFSQAQYLSRQEGTESTTDE----- 5256
Db 5125 RNHCTIITWAPLEDGSKITGYNEIREYGSTLTWVASDY-NVREPEFTVDKLRREFNDYE 5183
QY 5257 -----GOLPQVVEELRDLQVAPGTGLAKFQLKVGKGYPA 5289
Db 5184 FRVAINNAGKIPSLSPGPIKIQESGSRPQIVVKPED-TAQYNRAVFTCEAVGRPE 5242
QY 5290 PRLYFKDQGPLTASAHIRMTGKILHTLEIISVTRDSQGYAAYISNAMAAYSAARLL 5349
Db 5243 PTARLNRGRELPESSRYRFEASGVYKFTIKEWMDIDAGEYTVESNPGYSDTATANLV 5302
QY 5350 VRGDEPEEKASDVHQLVPPRMLERFTPKVKVKGSITFSVKVEGRPVTVHVL---R 5406
Db 5303 VCAPPVIEK-----DVPNTILP-----SGDLVRLKIYFSG-TAPFRHSLVLNR 5344
QY 5407 EEABERGVLWIGPDTFGYTVASSAQOHSILVLDVGRQHGQYVTCIASNAAGQALCSASLHV 5466
Db 5345 EE-----IDMDHPTIRIVEFDDHILITIPALSVREAGRYEYTVSNDSEATTFGFWLNV 5397
QY 5467 SGLPKVBEQKKEALISTFLOQTQQAISAOGLTASADLGGORKEEPLAAKEALGHS 5526
Db 5398 TGLPEAQ-----GPH 5409
QY 5527 LAEYGTBEFLQKTSQITEMVSAKITQAKLVPGGDSDESKTPSPRSGRSPSSIIQ 5586
Db 5410 ISNIGPSTATLSWRPPVTD-----GGSKITSYVV 5438
QY 5587 ESSSESDGDARGEIFDIYVVVADYLPGLAQDAITLREQOYVEVLDAAHPLAKLVRTKP 5646
Db 5439 EKRLSKD-----EWTVTGNVDMNYIVT--- 5463
QY 5647 TKSSFSRQGVSPAYLDERLKLSPWGNAAEPFGEAVSEDEYKARLSVIOQLLSSEQ 5706
Db 5464 -----GLFE-----NHEYEFVSA----- 5477
QY 5707 AFVEELOPLQSHHLQHLERCPCHPVIAVAGQKAVIFRNVDRIGRPHSSFLOEQCQDQD 5766

```

```

Db 5478 -----QNENGIG----- 5484
QY 5767 VAMCFIKQAQAEQYLEFLVGRVQAESVWVSTAIQEFYKKAEEALLAGDPSPQPPPLQ 5826
Db 5485 -----APLVS-----EHPITIALRPDPDPTSL- 5506
QY 5827 HYLEQVERVQRYQALLKELITENKARNQNCALLBQAYAVVYSALPQRAENKLHVSIMENY 5886
Db 5507 -----NLEIVQGGDYVTLWSOR----- 5524
QY 5887 PGTLEALGEPRIQGHFIVMEGAPGARMKGNHRHVFLEFRNHLVICKPRRDSRTTIVSYV 5946
Db 5525 -----PLSDG-----GGRLL-----RGYIV--EKQEEHDEWFR 5551
QY 5947 FRMMKLSIDLNQVEGDDRAFEVQWEREDSVRXYLLOARTAITIKSSWKEICIQORL 6006
Db 5552 NQNPSPNNYVNPILDG-----RKY--RYRVAYNDAGLSLAELODTL 5594
QY 6007 --ALPWRPDPPEBELADCTABLGTVTKLACRVTGTPKPVISWYKDGKAVQVDPHHILIE 6064
Db 5595 FOASGSGEGPKIVSPLSDLNEEVGRCVTFECEISGSPREYRNFKGCKELVDTSKTYLIN 5654
QY 6065 DPGSCALILDSTGVDSQYWCFAASAGNCSTLGLKILVQ-----VPPRFVNVKVRASP 6118
Db 5655 KGDKQ-VLIINDLTSDADDEVTCRATNSSGTRSTRANLRITKPRVFIPPKYHGGYEAQ- 5712
QY 6119 FVEGEDAQFTCTIEGAPYQIHWYKDGALLTTGNKFQTLSEPSGLLVIRAASKEDLG 6178
Db 5713 --KGETIELKIPYKAYPOGEARWTKDGEKIENNSKFSITDDK--FATLRTSNAERDYG 5768
QY 6179 LYECELVRLGRSARASAEIRIQSPMLQAEQCHREQLVAAVEDTLERADQEVTVLKL 6238
Db 5769 EYRVVSVSVSGSDGTAVTV- 5789
QY 6239 LGPKAPGSGTDLTGPGCPGCPALQETGSPVVTGTSEAPVPRVP--QPLLHEGP- 6295
Db 5790 -----ADVEE-PFRFPIENILDEAVI 5810
QY 6296 -EQEPEAI---ARACBWTVPJR-MEGAAMPAGAGTGELLMDVHSHVVRETQRTTYQOAI 6350
Db 5811 LSWKPPALDGGSLVNTYITEKEANGGWSW-----SPCAKSRITYTYTIE 5853
QY 6351 -THTARPSMQVTIEDVQAQT---GGTAQFAEAIISGDPQSPSVTWYKDSVOLVDSTRLSQ 6406
Db 5854 GLRAGKQYBFRIIAENKGGQSKPCPFTA--PVLIFGDERKRRRGYD---VDEQCKIVR 5906
QY 6407 QEGTYSVLVRHVASKDAGVYTCLAQNTGGQVLCRAELLVLGGDNEPDSEKQSHRRKLHS 6466
Db 5907 GKGTVSS-----NYDNYVFDIKQYY-----PQPEIKHDX---VLD 5940
QY 6467 FYEVKEEIGRGGVFGVFRVQHKGNKILCAAKFIPLRSTRQAQYR-ERDILAAALSHPLVT 6525
Db 5941 HYDIHEELGTGAFGVVHRVTERATGNFPAKFVMTPHESDKETVRKEITQMSVLARHPTLV 6000
QY 6526 GILDDOFEIRKTKLILILELCSSEELDRLY-RKGVVTEAEVKYIOQLVEGLHYLHSHGVL 6584
Db 6001 NLHDAPEDDNNVMVYEFMSGGELFEKVADEHNKMSDEAVEYMRQVCKGLCHMHENNV 6060
QY 6585 HLDIKPSNIMLVHPAREDIKICDFGFAQNTIPAEILQFSQYSPFVSPEIIOQNPFVSEAS 6644
Db 6061 HLDLKEPNIEMFTKRSNELKLDLFDGLTAHLDPKQSVKVTGTGTAEFAAEVAFEGKPVGYT 6120
QY 6645 DTWANGVSYLSLTCSPPAGESDRATLNVLEGRVSWSSPVAH--LSEDAKDIKATL 6702
Db 6121 DWMSGVLSYILLGSLSPFGGENDDETLRNVKS--CDWNMDDSAFSGISEDGDKDFIRKL 6178
QY 6703 QRAPQARPSAAOCLSHWPFLK-SMPAEAAHFINTKOLK 6739
Db 6179 LADPNTRMTIHOALEHPMLTPGNAPORDSQITPSSRYTK 6216

```

RESULT 9
A88852

protein unc-22 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
 C:Accession: A88852
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 283, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_elegans/
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 2103, 1999
 A:Accession: A88852
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-6831 <STO>
 A:Cross-references: GB:chr_IV; PIDN:CAA98081.1; PID:g3881830; GSPDB:GN00022
 C:Genetics: unc-22
 A:Map position: 4
 C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 3.9%; Score 1619; DB 2; Length 6831;
 Best Local Similarity 18.6%; Pred. No. 5.7e-40;
 Matches 1442; Conservative 933; Mismatches 2680; Indels 2712; Gaps 325;

QY 120 RVREGSBAATPCRVGSPRAVANSKDGRELBPDPGRVR-VBELGEA---SALRIRAAAR 175
 DB 7 KCKQG-DLLMECHLEADPOPTIAHQHSGNLL-EPGRVGVQTLTFLGGLYKATLVKEPN 64

QY 176 PRGGTYEVAENPLGAASAAALVDSDAADTASRPGTSTAALLAHLQRREMAEGA 235
 DB 65 AGDGAATCTARNOLGESNAININIFAGAGDEAKRGPSFV-----GK 108

QY 236 PASPPSTGTCTCTVTEGKHARLSCTVTGEPKPTVMKKDG-QLVTGRRHVVED-AQEN 293
 DB 109 PRIIPKDG-----GALLVMECKVKASATPVAKNMKGVLPSMGLYHAIFDLGDQT 160

QY 294 FV--LKLFLCKQSDRGLYTCTANLIGOTYSSVLVVRPAVPFKRLQDLVREXESAT 351
 DB 161 YLCOLEIRGSSDAGQYCNIRNDQOETNANALNPEEP-----DPSEQRERKST 212

QY 352 FLCEVPOPSTEAAMFKETRLMASAKYIE-BEGTERRTLVRNVSDADDAVICETPEGS 410
 DB 213 ---ASPRSSRGSGSRP-----SSPKSKMSRSGTPKR-----TLKPRGS 250

QY 411 RTVAELAVQGNLRLKLPRTAVRVGDTFMFCVELAVPVPVHMLNQEEVWAGRVAISA 470
 DB 251 PS-----KLRSTSTFV-----NEEV-----SOS 270

QY 471 EGRHTLTISQCLEDVGVAFAVAGDQCTSTRFCVAPRPFPLQPPVDPVWARMSSSVI 530
 DB 271 ESRSSRT-----DKXEVQVSG-----ASKRKP----- 294

QY 531 LSNSSPPHGRPVITDGLYVEKKLGYTWIRCHEAEWATPELTVDVAEENGFQFRVS 590
 DB 295 -DGLPPPG-----DEKKLR-----AGSP----- 312

QY 591 ALASFGOSPLYEPFGTVHAPKLAFTPLKAVQAVSGEVTFVSVDLTIVASAGFWLDGOA 650
 DB 313 ---STRKSP-----SRKASPTPSKGSAGG-----AASGTTGASASA 348

QY 651 LKASSVYIHCDRTRHTLTIREVPASLHGAKLFVANGIESSIRMEVRAAPGLTANKPPA 710
 DB 349 TSATSGGSASSDASRDYK-----RPP- 370

QY 711 AAREVLARLHEEAQLLAELSDDAAAVTWLKDGTLSPGPKYEQASAGRRVLLVRDVAR 770
 DB 371 -----IVLEASR 377

QY 771 DDAGLYECVSRGGRIAYQLSVQGLARFLHKDMAGSCVDVAVAGGPAQFECETSEAHVHVHW 830
 DB 378 -----SQTGRI-----GGSVVLEV-----QWQCHSSTI---IEM 403

QY 831 YKGMELGHSGERFLQEDVTRHRLVAAVTTRQDEGTYSGRV-----GEDSVDFLRVS 884

404 YRDGTLVRNSSE-YSQSPNGSIKLVNKLTEESGLYKCHAKCDYGEQSSAMVKLEQS 462
 QY 885 EPKVVFAKEQL-ARRKLOAEAGASATLSCEVAQAQTEVTWYKDGKLLSSSKVKCMEATGC 943
 DB 463 DVEBELMKHRDAEDYQKEBQKSOYL-----QAETKRVARRSKSKSKS----- 507

QY 944 TRRLVVQAGADAGEYSCAGGQRLSFHLDDVKEPKVFAKQVAH---SEVOAEAGANA 1000
 DB 508 -----PAPAKSKTTS-ESGRQEA---EVEHRSRSSVRPDPDEES 544

QY 1001 TSCVEAQAQAEVVMYKDGKLLSSSLKVHVKAECRRRLVVQAGKTADAGDYSCEARGQR 1060
 DB 545 QLD-EIP-----SSGLTIPEE---RRRELLGVGESDD----- 573

QY 1061 VSFRLHITEPKOMFAKESQSVHNEVOAEAGASAMLSCEVAQAQTEVTWYKDG----- 1111
 DB 574 -----EVSEISSELPFAGKPRKKTDDK 597

QY 1112 -KLLS-----SSSKVMEVKGCTRLRLVLPQAGKADAGEYSCAGGQVSFHLHITEPKGVF 1166
 DB 598 PKVSIAPVSTNKSSDDPESTPRR-----RSSIDMRRESVCEILEKTSTPLVPSGASGA 652

QY 1167 AKEOSVHNEVOAEAGATTAMLSCVA-OPQTEVTWYKDGKLLSSSKSVR---MEVKGCTRL 1223
 DB 653 PKIVEPENVTVVENETAILTKVSGSPAPTFRFGKSGREVISGGRFKHITDGKEHTVAL 712

QY 1224 VVQOVKADAGEYSCA---GGQVSFOLHITEPKAVFAKEQLVHNEVTEAGASATLS 1279
 DB 713 ALLKCRSODEGYTLTENVHGTDSADVLLVTSNGLDFRMLKHRE---SQAG----- 764

QY 1280 CEVAQAQTEVTWYKDGKLLSSSKSVRIEAGCWRQLVVQAGQADAGEYTCCEAGGQRLSP 1339
 DB 765 -----FQKDG-----EGGAGGG-----GGEKKPM 784

QY 1340 HLDVSEPKAVFAKEQ-----LAHRKVOAEAGAIATLSCEVAQAQTEVTWYKDGKLLS 1392
 DB 785 -TEAERQSLFPFGKWKNDIPLPEKTVQQQVDKICWKCTYSRPNAKIRYKDRKEIFS 843

QY 1393 SS---KVRMEAVGCTRLRLVVQAGADTGEYSCAGGQRLSFSLDVAEPKVFKAQOPVH 1449
 DB 844 GGLKYKIVIEKNVCT--LIINNPVDDTKYTCCEANGVPTHAQLTLVLEPKYSLNPLP 901

QY 1450 REVQAQAGASTLSCEVAQAQTEVTWYKDGKLLSFSK---VRMEAVG-CTRLVVQAG 1505
 DB 902 NTQIYRTKQAVLTCKVNTPRAPLVWYRGSKAQEGDPRIIEKDAVGRCT--LTIKEVE 959

QY 1506 QADAGEYSCAGSQRLS-FHLHVAEPKAVFAKEQAPASREVQAEAGTSATLSCEVAQAQTE 1564
 DB 960 EDDQAEWARTITQDVFQSVQVYVEEPRHTFVVPKSKQVNESDL---ATLETVDNDKDAE 1016

QY 1565 VTWYKDGKLLS-SSSKVMEAVGCTRLRLVQEAQADAGEYSCAGQRLSFLHVAEPK 1623
 DB 1017 VVWHDKRIDIDGVKFKVESNRKRLIINGARIEDHGEYKCTTKDDRTMAQL----- 1070

QY 1624 VVFAKEQ--PAHREVQAEAGASATLSCEVAQAQTEVTWYKDGKLLS--SSKVRVEAVGC 1679
 DB 1071 IVDKAKFEIYALKDTEVIEKDDVTLNMQTKTTPGIWFRNGKQISSMPGKKEQTSRNG 1130

QY 1680 TRRLVVQAGADAGEYSCAGGQRLSFLHVAELE--POISERPCR-----REPLVVKE 1732
 DB 1131 THTLKIGKIEMNEADVVEIDQAGLRGSCNVTLAEAKRPLINMKPKKIEAKGEPCEVVKV 1190

QY 1733 HEDILITATLATPSAATVTWLKQGVERRSKRH--ETASOGDTHLTIVHQAQVLDLSAIS 1790
 DB 1191 PFQIKGTRR-GDPKAI---LKNKGPIDBEKRLVVEIHKDDVAEIVFKNPQLADTKWA 1246

QY 1791 CRVG-----AEQDGF-----PVQVEEVAKFCR 1813
 DB 1247 LELGNSAGTALAPPFLFKDKPKPKPLETKNVTAEGLDLVWGTDPDDEGAPVKAYIIE 1306

QY 1814 LLE-----PVCGLGCTVTLACELSPACAEVVR-----CGNTQPRVKRFQWVAEGP 1861

Db 1307 MQEGRSNWAKVGETKGTDFKVKDLKEH-GEYKFRVKALNECGLSDPLTGE--SVLAKNP 1363
 Qy 1862 VRSUTVLGLRAEDAGEVCSRRDDHTSAQLTVSVPRVVKFMSGLSTVVAEGEATFCQV 1921
 Db 1364 -----YGVPGKPKNDIAIDVDKDHCT-----LWPEPDEGGAPI--- 1398
 Qy 1922 VSPSDVAVWFRDGLALOPSEK-----FAISQSGASHSLTISDLVLEDAGQI-TVAEAGA 1975
 Db 1399 -----TGIIIREREKESEKWHQVQTKPDCCELTDKKVVEDKEYLYRVXAVNK 1446
 Qy 1976 SSSAALVRBAPVLFK-KKLEPQ-----TVEERSSTVLEVELT-RPWPELRWTRNA 2024
 Db 1447 AGPGDPCDHGKPIKMKAKKASPEFTGGIKDLRLKVGETIKYDVPISGPEPLPCLMWVNG 1506
 Qy 2025 TALAPGKVIHAEGARHLVLHNVGPDAGFF-----CCEPDDKTKQAKLVEMR 2075
 Db 1507 KPLKAVGRVMSRGRGHKHIKINAVRADSGKTTILKNSGGSCD-----STATVUGR 1561
 Qy 2076 QVRLVRGLQAVEAREQGTATME-----VOLSHADVDGSWTRDGLRFQOG 2119
 Db 1562 PTPKPGPLDIADVADG-ATLSWNPDDDDGGDPLTGYIVAEQAQMDNKGKIEVG----- 1614
 Qy 2120 PTCHLAVRGPWHTLTSLGAPEDSGLMVFKAEGVHTSARLVVITELPVSRRPQ-QDVVIT 2178
 Db 1615 -----KVDPNNTTLKVNGLR--NKGNYKFRVKA VNEGE-----SEPLASOQYTO 1657
 Qy 2179 EKEKVTLECELSRP-----NVDVWV-----LKDG-----VELRAGKTMIAAQQGAC 2219
 Db 1658 IKOPWDSPGKGRPEITDFADRIDIAWPPHKGDDGAPIEYIVVDRDPDK----- 1709
 Qy 2220 RSLTIYCEPADQGVYCDADHQAQSSASVKVQGRYTLIYRVLADAGIEQIVVAENAS 2279
 Db 1710 -----EWKE-----VVRPDTNASTISGLKEKEYQFRVAVNKAQPGQ-----PSEPS 1752
 Qy 2280 RAQL-RVKELPVTILVRPLDKIAKEKRGVLEQCVSRASA-----QVRWFKG 2325
 Db 1753 EKQAKPKFIPAWL-----KDNLKS-TVKAGATVRWEVKIGGPIEYKFKG 1801
 Qy 2326 SQELQPGKVELVSD-----GLYRKLIIISDVHAED-----DT 2358
 Db 1802 NOQLENG-IQLTIDTRKNEHTILCIPSAMRSDVGEYR-LTVKXSHGADBEKANLATVLD 1858
 Qy 2359 YTCADGVKTSQAQFVEZEQISITVIRGLQDVVNEP-----APAWFEC----- 2400
 Db 1859 PSKPNGLVSDVF---EDNLNLSWKPPDDGGEPYBYEVEKLDIATGRWVPCAKVOT 1915
 Qy 2401 -----ETSIPIV----- 2407
 Db 1916 KAHIDGLKKGQYQFRVAVNKEGASDALSTDKDTAKNPYDEPGKTGTPDVVDWADRV 1975
 Qy 2408 -----RPPKWLIGKTVLQ-----AGNVGLB-----QEGTVHRLMLR 2440
 Db 1976 SLEWPPKSDGGAPITQYVIEKKGKGRDWQECQKVSQDQTNABILGLKEGEBYQFRVKA 2035
 Qy 2441 TCSMTGPVHFTVCKSRSSARLVSDIPVVLTRPLEPKTRELQSVVL-----SCDF--- 2492
 Db 2036 V--NKAGP-----GEADSPKSVAK-----PRLKFWIDREAMKITIKVGNVDFDVP 2083
 Qy 2493 ---RPAPKAVQYKDDTPLSPSEKFKMSLEQMAELRILMLPADAGVYRQQA-----GSA 2545
 Db 2084 VRGEPKPKK-EWIFNEKPVD-DQKIRIESDYKTRFVLRGATRGHAGLTLTATNASGSD 2141
 Qy 2546 HSSTEVVAREVITVTPLODAEATEBGWASFSCELSHEDEVEWLSL-----NGMPLNDS 2601
 Db 2142 KHSVEVILGKPPSPPLGPLEVSNVYE-----DRADLEWKPVEDDGGAPI---D 2186
 Qy 2602 FHEISHK-----GRRHTLVKGIQRADAGIVRASSILKYSTSARLEVRVYKPVVFLK 2651
 Db 2187 HYEIEKMDLATGRWVPCGRSET-----TKTTPNLPQHGHEYKFRVAVNKEG 2233
 Qy 2652 ALDDLISAE-----ERGTIALQCEVSDPEAHVVRKDGVLQ----- 2687
 Db 2234 ESDPFTTNTAILAKNPYEPVPGVKDKPEL-VWDKDXHVDLAWNAPDDGGAPIEAFVIEK 2292

Qy 2688 -----PSD-----KYDFLHTAGTRGLVVHVDVSPEDAGLYTCHVGSE 2723
 Db 2293 KNGRWEELVVPDQOKTATVNLKEGEBYQFRIGARNKAGTGDPSPDSD-----RVVAK 2346
 Qy 2724 ETRAVRVHDLHVGITKELKTMVELEGESCFEVLSHESASDPAMTVVGKTVGSS--- 2780
 Db 2347 PENLAPRIH-----REDLSDTTVKVGATLKFVHIIDGEPAPD-VTWFSNGKIGESKAQ 2399
 Qy 2781 -----SRF-----QATQGRKYLIVREAAAPSDAGEVVFVSRGLTSK-----ASLIVR 2823
 Db 2400 IENEYISRFPALPKALRKQSGKYTTATNINGSVITINIKVSKPTKPKGPIEVDVFE 2459
 Qy 2824 ERPAAIKPLEDQ-----WV----- 2838
 Db 2460 DBATLDWKPDEDDGGEPIEFYIEKMTKDGWPCGRSGDHTFTVDSLNKGDHYKFRVK 2519
 Qy 2839 -----APGSDVELRCEL-----SRAGTP-----VHW-----LKORKA----- 2865
 Db 2520 AVNSGSPDPLETETEDILAKNPFDRPGRPEPTDMDSDHVDLKWDPPLSDGGAPIBEY 2579
 Qy 2866 -IRXSOKYD-----VVCETMAMLVIRGASLKDAGEYTCV-----EASKSTA 2907
 Db 2580 QIEKTKYGRNEPAITVFGQTTATV--PDLTPEEYEFVRVAVNKGGPSDFSDASKAVI 2637
 Qy 2908 SLHVEEKANCFTTELTN-QVBEKGTAVFTCKTE-HPAATVTWR-KGELLEASGHKQPSQ 2965
 Db 2638 AKPRNLKPHIDRALKNTIKAGOSISFDVPVSGEPAPTIVTHWPDNREIRNGRVLKN 2697
 Qy 2966 EGLTURLTISALEKADSDTYTCIDGQAQSRQALLVQGRVHLIE-----DLEDVD- 3015
 Db 2698 PEYQSKLVVQMERGDSGTFTIKAVNANGEDEATV---KINVIDKPTSPNGPLDVSVDHG 2754
 Qy 3016 -----VQEGSAT-----FRCR 3027
 Db 2755 DHVTLNWRAPDDGGIPIENYVIEKYDTASGRVPAKAVAGKTTAVVDGLIPGHEYKFR 2814
 Qy 3028 ISPANYE---PVHWF---LKTPL-----HANELNEID-----AQPG----- 3059
 Db 2815 VAAVNAEGESDPLETFTGTLAKDPFKPKTNAPEITDMDKDHVLEWKPPANDGGAPIE 2874
 Qy 3060 -----YHVLTLRQLA---LKDSGTIYF-----EAG-----DORASAA 3088
 Db 2875 EYVEMKDEFSPFNDVAHVPAQNTNAVNLKSGSKYEFRIACNKAAGLDPSPSASAV 2934
 Qy 3089 LATVEKPSVFR-ELTDATITEGEDTLVCTETSCDIP-MCWTKDGKTLRSARCOLSHE 3146
 Db 2935 AKARNVPPVIDRNSIQEIKVKAGQDFSLNIPVSGEPTTITWTFEGTVPESDDRMKLNE 2994
 Qy 3147 GHRAQLLITGATLQDSGY--KCEAGGACSSSIVVHA-----RPVRFQALKDLEV 3196
 Db 2995 DGKTFHVARLRSDTGYIIEKENENGTDIAEVKTVLDHPSSPRGFLDVTNIVKDGCD 3054
 Qy 3197 L-----EGGAATLRCVLSSV-AAPVKW-----CYGNN-----VLR 3225
 Db 3055 LAWKEPEDDGAIEI SHYVIEKQDAATGRWTACGESKDTNFHVDDLTQGHEYKFRKAVNR 3114
 Qy 3226 PGDKYSLRQEGAML-----ELV-----VRNLRPODSGRYSCSGDQTT 3263
 Db 3115 HGDSDPLEAREAITAXDPFDRADKFGTPEIVDWDKDHADLKWTPPADGGAPIE-GYLVE 3173
 Qy 3264 SATLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAPVEWRKGSETLRDGRYCLRDQ 3323
 Db 3174 XRTPSGDWVPATVG-----AGE-LTATV-----DGLKPGQTY----- 3205
 Qy 3324 GAMCELOIRGLAMVDAAYEYSCVCGEERTSASLTIRMPAHFTGRURH-----QESIEGAT 3378
 Db 3206 -----QFRVKALNAGE-----STPSPSRSTMVAKPRHLAPKIRNDRMFVQVRKAGOT 3253
 Qy 3379 AT--LRCELSKAAPVEW-RKGRESLRDGRHSLRQDGA VCELOICGLAVADAGEYSCVC- 3434
 Db 3254 LNFQVNEGEPAPKLEWFLNGSLSSGNNTHIDNNTDNTKLTITKSTARADSGKXIYAT 3313

Db 5128 LEDGSGKITGYNVEIREYVSTLWTVASDY-NVRBPFTVDKLRBNFNDYFRVAINAAGK 5186
 QY 5257 -----GOLPQVBEIRDLOVAPGTRLAKFQKLVKGYAPAPRLYWFQDQP 5300
 Db 5187 GIPSLPSGPIKIBESGSRPQIVKPED-TAOPYNRRRAVFTCEAVGRPEPTARWLNRGE 5245
 QY 5301 LTASHIRMTGKILHTLEIISVTRDSGOYAAIISNAGYASABLLVGRPPPEPKP 5360
 Db 5246 LPESRVRFEASDGYKFTIKEDWDIDAGEYTVESNPNYSGDTATANLVQVAPVIBK-- 5303
 QY 5361 ASDVHEQLVPPMLERFPPKVKSGSSITFSVKVEGRVPVTHWL---REAEGRVLWIG 5417
 Db 5304 --DVPNTILP-----SGDLVRLXIYFSG-TAPFRHSLVLNREE-----ID 5340
 QY 5418 PDTPOYTVASSAQHSLVLLDVGRHQOQTYTCIASNAGQALCASLHVSGLPKVBOEK 5477
 Db 5341 MDHPTIRIVFDDHILITIPALSREAGRYEYTVSNDSGEATTFWNLVTGLPEAPQ--- 5397
 QY 5478 VKEALISFLQGTQCAISAQGLTASPADLQGRKEBPLAAKALHLSLAEVCTEEFLQ 5637
 Db 5398 -----GPHISNIGPSTATL 5412
 QY 5538 KLTGQITMVSAKITQAKLVPGGSDSDSKTPASPRHGRSRPSSISQSSSSSDGDA 5597
 Db 5413 SWRPVTD-----GGSKITSYVVEKRLDKD--- 5438
 QY 5598 RGEIFDIYVTVADYILPLGABQDAITLREGQVEVLDAAHPLRLWLVRKPTKSSPSRQGW 5657
 Db 5439 -----EWVTYTSNVKMNIVT----- 5455
 QY 5658 SPAYLDRRLKLSPEWGAAPFEGEAVSEDEYKARLSSVTOELLSEQAFVEBLQFQS 5717
 Db 5456 -----GLFE-----NHEYEFVSA----- 5469
 QY 5718 HHLQHLERCPHVPIAVAGCKAVIPRNVDRDIGRFHSSFLQELQCCDQDDVAMCFIKNQA 5777
 Db 5470 -----QENGIG----- 5476
 QY 5778 FEQYLEFLVGRVQABSVVSTAIOBFYKKAEEALLAGDPQPPPPPLQHVLEQVVERVQ 5837
 Db 5477 -----EPIIARLPDPTSP----- 5498
 QY 5838 RYQALLKELIRKARNRQNCALLEQAYAVVSGALPORAENKHLVSLMENYPTLEALGEPI 5897
 Db 5499 -----NLEIVQGGDYVTLWSOR-----PL 5518
 QY 5898 RQGHFIVWEGAPGAMPKGNHRHVFRLNHLVICKPRDSRTDTVSYVFRNMKLSID 5957
 Db 5519 SDG-----GGRL-----RGIV-----EKQEEHDEWFRCNQPPPNYN 5554
 QY 5958 LNDQVEGDDRAFEVWQEREDSVRYKYLQARTAIKSSWVKEICGIIQRL--ALPVMRPPD 6015
 Db 5555 VNLIDG-----RKY--RYRVFAVNDAGLSDLAELDQTLFQASGSGEPK 5597
 QY 6016 FEEELADCTAELGETVVKLACRVGTGPKVI SWYKDGKAVQVDPHILIEDPGSCALILD 6075
 Db 5598 IVSPLSLDNBEYGRVCTFECEISGSPREYRWFKCKELVDTSKYTLINKGDKQ-VLIIN 5656
 QY 6076 SUTGDSGOYCMFAASAAGNCSTLGIIVQ-----VPPRFVNVKVRASPFVGEDAOPTC 6129
 Db 5657 DLTSDDADEYTCATNSSGSTRANLRIKTPRVFIIPKYHGGVEAQ--KGTEILKI 5713
 QY 6130 TIEGAPYPOIRWYKDGALLTGNKFTLSRPSGLLVIRAAASKEDIGLCEBVLNRLG 6189
 Db 5714 PYKAYPQGEARWTKDGEKTEENSKFSITDDK--FATLRISNASEDYGEYRVVNSVG 5771
 QY 6190 SARASAEILRIQSPMLQAOEQCHREQLVAIVEDTLERADQEVTSVLKRLGPKAPGSTG 6249
 Db 5772 SDSGTUNVT----- 5781
 QY 6250 DLTFGPCPCRGAPALQETGSPVVTGTSEAPVPRVP--QPLHLEGP--QOEPAI--- 6302
 Db 5782 -----ADVPE-PPRFPIIENILDEAVILSWKPPALDGG 5813

QY 6303 ARAQEWTVPIR-MEGAAMPFAGCTGELLWDVHSHVVRVRETTORTTYTQOAI--THTARPPSMQ 6360
 Db 5814 SLVTNTYTIKREKAMGWS-----SPCAKSRYYTYTTIEGLRAGKQYBFR 5856
 QY 6361 VTIEDVQOAT--GGTAQFEALIEGDPQPSVTVYKDSVQLVDSTRLSQOQEGTYSVLVR 6417
 Db 5857 ITAENKGGGSKPCEPTA--PVLI PGDERKRRGYD-----VDEQKIVAGKGTVSS---- 5905
 QY 6418 HVASKDAGYVTCIAQNTGGQVILCKAELLVLGDNEDPDSKQSHRRKLHSHFYEVKEEIGRG 6477
 Db 5906 ---NYDNVYVDFIKWQYV-----PQVEIKHDE--VLDHYDIHEELGTG 5943
 QY 6478 VFCFVKYVOHKNGKILCAAKFIPLRSRTAQAYR-PRDILAAALSHPLVTGLDQETRTK 6536
 Db 5944 AFQVHRVTERATGNNFAAKFVMTPHESDKETVRKEIQTMSVLRPTLVNLDHAFEDDNE 6003
 QY 6537 LILILSCSSEELDLRY-RKGWVTEAEVAVKVIQQLVEGLHYLHSHGLVHLDIKPSNIML 6595
 Db 6004 MWIYEFMSGGELFEKVADEHNKMSDEAVEYMRQCKGLCHMHENNYVHLDLKPENIMF 6063
 QY 6596 VHPAREDIKIDCGFAQNTIPPAELQPSQYSGSFVSPELIQONPVSEASDIWAMGVISL 6655
 Db 6064 TTKRSNELKIDFGLFAHLDPKQSVKVTGTGTAFAAEVAEGKPGVGYTDMMSVGVLSVI 6123
 QY 6656 SLTCSPPFAGESDRATLLNVLEGRVSWSSPMAAH--LSEDAKDFIKATLQRAPOAPPSAA 6713
 Db 6124 LLSGLSPFGGENDDETLNVKS--CDNMWDDSAFSGISEDGKDFIRKULLADPENTMTIH 6181
 QY 6714 QCLSHPWFLK-SMPABEAHPINTKQLK 6739
 Db 6182 QALEHFWLTPGNAPGRDSQIPSSRYTK 6208

RESULT 10
 T13931
 Projectin - fruit fly (Drosophila melanogaster) (fragment)
 C:Species: Drosophila melanogaster
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #ext_change 17-Nov-2000
 C:Accession: T13931
 R:Daley, J.; Southgate, R.; Ayme-Southgate, A.
 J. Mol. Biol. 279, 201-210, 1998
 A>Title: Structure of the Drosophila projectin protein: isoforms and implication for pr
 A:Reference number: Z17815; MUID:98300339; PMID:9636710
 A:Accession: T13931
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-6658 <DAL>
 A:Cross-references: EMBL:AF047475; NID:g3337430; PID:g3337431; PIDN:AAC27550.1
 C:Genetics:
 A:Gene: projectin
 A:Cross-references: FlyBase:FBgn0005666
 A:Map position: 4
 A>Note: intron positions not resolved (incomplete sequence)
 C:Keywords: muscle

Query Match 3.4%; Score 1392; DB 2; Length 6658;
 Best Local Similarity 18.4%; Pred. No. 2.7e-33;
 Matches 1567; Conservative 1067; Mismatches 2994; Indels 2902; Gaps 382;
 QY 182 YEVRANPLGAASAAALVVDSDAATASRPGTSTAALLAHLORREARAEAGAPASPPS 241
 Db 2 FRVANVRAG-----PGEPSDKTSIIAKCFVKPFVIGE----- 36
 QY 242 TGTRTCTVTGKHARLSVYVGPETVWKGQQLVTEGRRHVYVEDAQENFVLKILFC 301
 Db 37 -GLKNVTVKGQITRFDIKYDGEPEPAATVWGTDLNLFKDNCRICLDQLERNSSITIKS 95
 QY 302 KQSDRGLYTCTASNLVGQYSSVLVVRE----PAYPEKKRLQDLEVRKESATFLCEVP 357
 Db 96 VRKDTGKYLVLNSSGTTESEAQVVLDRPLPPGGPFPE-----EIR 139
 QY 358 QPSTEAAWFKETRLWASAKYIGIEERGTERRLTVRNVSDDDAVY-CETPEGSRVVAELA 417

| | | | | | | |
|----|------|--|---|--------------------------------------|-------------------------|------|
| QY | 2310 | ECQVSRASAOVRWFKSGQEL----- | QPEPKYELVSDGLYKRLIISD | 2350 | | |
| Db | 1876 | TVPIISK-----RPHSGKYTLKAENESGRDEASFVIVLDKPCP----- | PEGPLR-----VTD | 1922 | | |
| QY | 2351 | VHADEDEDTYTC----- | DAGDKVTSQAQFFVEBQSIITVIRGLQDVTWMEAPAWFECET | 2402 | | |
| Db | 1923 | VHKEG----- | CKUKMAPDDGGLPID-HYIIEKXDVESGRWL----- | PSGRFKGSFA | 1969 | |
| QY | 2403 | SIPSVRPPKWLGLKTVLQAGNGVGLQEQTGVRHLMRLRRTCTMTGPHFTVVGKRSARL | 2462 | | | |
| Db | 1970 | ELNLEPS----- | HEYKFRVLAVN----- | TEGES----- | 1993 | |
| QY | 2463 | VNSDIPVILTRPLEPKTGBELQSVLVSCDF----- | RP-APKAVOWYKDDTPL----- | SPSEKF | 2514 | |
| Db | 1994 | ----- | EPLTGB-----QSVIAKNPDPFGPKGTFPEAVDMDKHVCLVWMPPIINDG | 2038 | | |
| QY | 2515 | KMSLEGQMAELRILRLMPADAGVYRCQAGSAHSSTEVTVAREVTVTGPLQDABATEEGW | 2574 | | | |
| Db | 2039 | GSPIYGVVYVKR----- | ----- | EKGTDKWKIGTEITIPCLGECKAT----- | 2075 | |
| QY | 2575 | ASPSCELSHEDBEVWSLNGMPLYNDSPHFHISHKGRHTLVLSIQRADAGIVRASSLUKV | 2634 | | | |
| Db | 2076 | ----- | ----- | VPTLNEC----- | EYEFKVKAINAAGPEPSDASKPI | 2107 |
| QY | 2635 | STSGARLEVRKVPVF--LKALDGLAEBRGTLALOCEVS-DPEAHVWVRKDGVQLGPSDK | 2691 | | | |
| Db | 2108 | ITKPR--KLAFTLDPTKNIITYFKSGEPIFLINISGEPADVTWQNNKSVQVITSP | 2164 | | | |
| QY | 2692 | YDFLHTAGTGLVYHVDSPEDAGLY--TCH--VGSEETRARVRHDLHVGITKRL--T | 2744 | | | |
| Db | 2165 | SHIENLPYNTKYINNPNPERKDTGLYKISAHNFYQDQVEFOINI----- | ITKPKGPGP | 2218 | | |
| QY | 2745 | MEVLE--GBSCSF----- | ----- | ECVLSHESASDPAMWTVGKTVGSSSRFOATRQG | 2789 | |
| Db | 2219 | LEVSEVHKOGCKLWKWKXKDDGGBPVESYLVEKFPDPTGIMLPVGRSGPYNVVDGLVPG | 2278 | | | |
| QY | 2790 | RKYILVREAAPSDAGEVSVFVSRGLTSKASLIVRERPAALIKPLEDQWVAPGEDVELRCE | 2849 | | | |
| Db | 2279 | HDYKFRKAVANKGESEPLETLGSLIAKDPPFSVPTKPGV--PEPTDWTANKVELAWPEP | 2335 | | | |
| QY | 2850 | LSRAGTVPVHWDKRAIKRSQXVDVCEGTM----- | AMLVIRGASLKADAGEYICEVEAS | 2903 | | |
| Db | 2336 | ASDGGSPIOGY----- | IWEVKDYSLPWEKALETNSPTTATVQG--LIEGNEYQFRYVAL | 2389 | | |
| QY | 2904 | KSTASLHVBEKANCFT----- | ----- | EELNLOVEBKGTAVFTCK--TEHPAATVWR | 2949 | |
| Db | 2390 | NKGNLSEPDSPSKIFTAKKRYIAPKIDRNLNLI--SSGTALKLDANITQOPAPKVEW- | 2447 | | | |
| QY | 2950 | KGLELRASGHQPSQEGITL----- | RLTISALEKADSTYTCDIGQASRAQLLVQG | 3002 | | |
| Db | 2448 | ----- | KLNSYHLQSGKNTIETPDYTKLVIRPPQRTDSGEYLVTANTSGKDSVLVN- | 2500 | | |
| QY | 3003 | RVVHIED----- | LEDVDV--QEGSSATFCRISPAWYEPVHMF-LDK----- | TP | 3044 | |
| Db | 2501 | ----- | AVITDKPSPPNGPLQSDVHKGCHULWK-K-RFSGHGGTPIEFYQDKLEBPTGWCWIP | 2556 | | |
| QY | 3045 | LHANELNEIDA---QPGGYHVLTLROLAKDSGTYTFEAGDQORASALURVTEKPSVFSRE | 3101 | | | |
| Db | 2557 | SCRSTAPQVDVTCGLSPGNEY--KFRVSANVASGESQPLVGDBSIVARNPFDEPKPENLK | 2614 | | | |
| QY | 3102 | LTD----- | ----- | ATIIEGEDLTVCTSTCDIPMCWTDKCKTLRG----- | SARCQLS- | 3144 |
| Db | 2615 | ATDWDKHDVLAWTPVPIVGGSPi----- | ----- | SCYIEKODKYGKWRALDVPADOCKATI | 2667 | |
| QY | 3145 | ----- | HEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRVHAR-PVRFOEALKDLVELEGG | 3200 | | |
| Db | 2668 | PDVLEGQTYKFRVSANNAAGTG---EPDSPTPPIAKARNKPPIIDRSSLIVEVRIKAGQ | 2723 | | | |
| QY | 3201 | AATLRCVLSVAAP-VWVCYGNVNLAPGDKYSLROEGAMLELVLRNLRPQDSGRYSC--- | 3256 | | | |
| Db | 2724 | SFTFDCKVSGEPAPQTKLLIKKEVYSKDNVNTVNDVNTKLVNSATRSDSGIYTVFAE | 2783 | | | |

Db 3574 --SPSEPSDMMRPRYPKPIITPLNEVRIKGLIFHTDIHFGEPAEATWLNPNP 3631
 QY 4290 LQANEMNDITVEQGLTHLLTKHKTLEDAGTVSFHVGTCSEAEQKLVKTAKNTVVRGLENV 4349
 Db 3632 LLSNDRSIT--SIGHUSVHTVNCORSDSGITLH-----LUNSSGI 3672
 QY 4350 BALEGGEALF-----ECOLSQPEVAHAHTWLLDDPEFVRTSENAEYVFFENGRLHLLL 4401
 Db 3673 D--EGSFELVLDPRGPEGMEYEETANSVTISWKP----- 3708
 QY 4402 KNLPRDSCRVTFLAGDWTSAFTVR-----CWRLEILEPLKNAAVRAGAQAARFCTTL 4455
 Db 3709 ---PKON-----GGSEISSYVIEKRDLTTHGGW----- 3733
 QY 4456 SEAVP--VGEASWYINGAAVQDDSDWTVTADGSHQALLLSAQPHHAGEVTFACRAVAS 4514
 Db 3734 ---VPANVYSAKYNHAVPR-----LLEGTYELRVMAENLOGRSDPLTSDQPVAK 3783
 QY 4515 ARLTVGLPDPPEDAEVVAHSHHTVTLSWAAPMSDGGGLCGYRVVEKAGATQW-RLCH 4573
 Db 3784 SCYTVPGAGPKP---ELTSDKNHITIKWKQPIGNGGSPGIIGYDIERDVTGWIKING 3840
 QY 4574 ELVPGPECVGLAPGETYRFRVAAGVPGAGEPVHLPQTVRLAEPKPPVPPOPSAPESR 4633
 Db 3841 QPVTAEQDDRVTNSHOYQYRISAVNAGN-----KTSEPSAIFNARPLEKPR 3891
 QY 4634 -----QVAGEDVSLELV--ABAGEVIWHKMERIQGGFVVSQGRQOMLV 4681
 Db 3892 FYFDGLIKRIKVRAGEPVNLNIPISGAPTIEIWKRGDLKLEBKRSISETNSERTLFR 3951
 QY 4682 IKGFTAEPQGVHCHGLAQSGICPAANTFOVALSPASVDEAPQSLPPEAAQEGDLHLLWE 4741
 Db 3952 IDDSNRDSGKI-----TVTANFEGKTDADIEVIVVDKPS-PP-----EGPLSY--- 3995
 QY 4742 ALAKRRMSREPTLDSIS---ELPEEDGRS-----QRLP---QBAEEV 4778
 Db 3996 -----TETAPDHISLHWYSFKDDGGSDITGYIIEFTFEGVDWKVPVPGTCPTNFT 4046
 QY 4779 APDLSEG-----YSTABELA-----RTGDADLSH 4802
 Db 4047 VKNLVEGKXYFRIRAEENYIGASEALEGKPVLVKSPFPDPCAPSQPTISAYTPNSANLEW 4106
 QY 4803 TSSDD-----ESRAG-----TPS-----LVTYLLKAGRPCTSPASKVGA 4837
 Db 4107 HPPDDCGKPGITGYIVERREGGEWIKCNNYPTENTSYTVSNLELDGARFEFRVLAVNEAG 4166
 QY 4838 PAASVVKPQOOQELAA-----VRPPLGLDSTKDLGDSMDKAAVKIQAFKQYKV 4888
 Db 4167 PGHPS-KP--SDPMTAEHCRYRDPPEPPKPDRIITRNGVTLNWRPRTDGKRIKGYV 4222
 QY 4889 RKENKQOQB-----PMPSHF----- 4904
 Db 4223 --EVRPKNGKDWKTVDIPINSTYTVPSLKEGEYSFRVVAENEVGRSDPSKPSQPIITI 4280
 QY 4905 -----GDTEAQVDALRLCEVNVASKADVRARWLKXGVELTDPGRHHHIDQLG 4950
 Db 4281 BEQPNKPCWELGKVRDIVCRAGDDFSIHVPYLAFFKPNFAWYNDNMLDDNNRVH-KHLT 4339
 QY 4951 DGTCSLLIAGLDADAGCYTQVSNKEGQVTHSACVVVSGSEASESSGELDDAFRA 5010
 Db 4340 DDAASVVVNVNKRSDSGYRLQFKDTSG-----FDTA 4371
 QY 5011 ARRLHRLFRKSPAEV-----SDELF-----SADEGPA-----EPEEPADWQTYRED 5054
 Db 4372 TINRVLDPRPPTRLRADVFSGLTLYNPNDDGSAIQNYVIEKKB-ARSSTWSKV 4430
 QY 5055 EHFICIRFEALTEARQAVTRFOEMFATLIGVIELVQEQPRRVECMISKETP--APVVP 5112
 Db 4431 SSFCTVLIF---VRIRNLVNLKEYDFRVI---AENKYGQSDPANTSEPIARHFPDIPNTP 4484
 QY 5113 PEPLPSLLTDAAPVFLTELQNBQVQDGYPVSFVCTVVGQPMPSVRWFQDKGLEEDDHY 5172

Db 4485 GIPHGIDSTEDSITIATWTKPKH-----DG-----GSPI-----TGYIIEK---R 4520
 QY 5173 MINEQOQGHQHLIIITAVVPADMVVRCLAEENSMGVSSTKAELRVDLTSTDYDTAADATES 5232
 Db 4521 LLSDDKW-----TKAVHALCPDLSCKIPNL--IENAEYEFV-----AAVNAAGQ 4563
 QY 5233 SYFSAQGYLSSREGEGETESTDEGQLPQVVEELRDQVAPGTRLAKPOLKVKGY--PAP 5290
 Db 4564 SAYSGSSDLIFCRPPHAPKITSLSIP-----DMTVIAGD---EPRITVPYHASRP 4613
 QY 5291 RLYWFKDQOPLTASAHIMTKKILHTLEIISVTREDSGOYAAAYISNAGMAGYASARLLV 5350
 Db 4614 TASWSLNGLEVIPGBRIKFDNSDYASMYNKSAGKDETSYITILNNKGSDDTASCHTV 4673
 QY 5351 RGPDEPEEKP--ASDVHEQLVPPRMLERFTPKVKKSSITFSVKVEGRPVPTVHMLREE 5408
 Db 4674 VDRPLPPQGPLNAYDI-----TPDT-----CTLAWKTPDLDGSGSPITNYVVEK 4716
 QY 5409 AERGVLWIGPTPGTYTVASSAQOHSVLVLDVGRCHQGTCTCIASNAAGQALCSASLHV-- 5466
 Db 4717 LNSGSMV-----KISSFVNTHYDVMGLEPHYKYNFRVRAENQYG---LSDPDLIE 4766
 QY 5467 -----SGLPKVEBOEKVKEALIST-----FLOG-----TT 5491
 Db 4767 PMVAKHQFTVPDEPQPKVIDWDSGNVTLIWTRPLSDGSGRIQGYQIEYRDILNDSMNA 4826
 QY 5492 QAIASQGLSETASFADLQORKEEPLAAKEALGHLSLAEVGTTEEFLQKLTQSITEVMSAKI 5551
 Db 4827 YDIYIKGTQYLYNLINGSYEYFRIKAKNAAG---LSKPSPSLRFKL----- 4871
 QY 5552 TQAKLOVPGGSDSDSKTPSPASRGRSRPSSSIQSSSESEDGDAR--GSIPO----- 5603
 Db 4872 -KGKFTVP-----SPPGAPQVTRVGNVYDLKWEKPLRDGSRITGVIIERRDIG 4921
 QY 5604 -IYVVTADYPLGABQDAITIRE-GQY-----VEVLDAHP 5637
 Db 4922 AVWKNCNDYVLDTEYTMNLIEMGDYEFYFVAVNSAGRSEPSLCTWPIKVCEVLGGKKP 4981
 QY 5638 LEWLVRTKTPSSPSRQWSPAYLDREL-----KLSP--EW--GAEEAPEFGEAVSE 5687
 Db 4982 -DWITRL-----QKVAPFGKDYTLQCAASGKSPFTARWLKNGKBIQWNGRMTCD 5031
 QY 5688 DB---YKARLSSV-----IQELLSSEQAFVEELOFQSHHLQHLERCP----- 5727
 Db 5032 SKDGVFRLHISNVQDGDYTCAMNS-LGFVNTSGYLKIGS--PIINRCPSELKLPEDG 5089
 QY 5728 --HVPPIAVAGOK--AVIFRNVDRIGRHFSSFLQELQOQDCTDDVAMCFIKQAQAFQYLE 5763
 Db 5090 NSKIKIFYSGDQPLTVILKKNNEV-----ISDSNDTTHV-----KNVIFDDYVA 5133
 QY 5784 FLV-----GRVQAESVWVS-TAIOEFYKKAEEALLAGDPSQPPPP----- 5823
 Db 5134 IVIANIVKSDGPGYQIEFTNESGATGEFY-----VHITGMPAPTGMGSIYINKNSC 5167
 QY 5824 -----PLOHYLEQPERVQRYQALLKELRNKARNQNCALLEQAQAVVSAL 5870
 Db 5188 MLNMRPPYDGLKVSHYVTERKDVSSPHWITVSTCKDTAFNVQ--GLIENQBYIFRVM 5245
 QY 5871 PQAENKLVHSLMENYPG-TLEALGEPIRQGHFTVWEGAPCARPFWKGNHREVLFRNHL 5929
 Db 5246 -----AVNENGMGPPLGL-NPIRAKDPIDPSPPAV-----LAKSLREE 5284
 QY 5930 VICKP-----RRSRTDTVSYFRNMMLKSLIDLDNOVGDDRAFE 5970
 Db 5285 TLCPNPEWKPSDGAHTQGYWIDKREVGSKHMAVGNATICAQINCINLIEGRQYEF 5344
 QY 5971 VWOEREDSVRYLLQARTAIKSSWKEICGIIQORLALPVWRPDPFEEL-ADCTAEIGE 6029
 Db 5345 IFAQN-----VAGLSIAIVR-----FTSQNRSTGGFPASIDCTATHANCIO--NH 5389
 QY 6030 TVKLACRVGTGPKPVISWYKDGKAVQVDPHHILIEDPDG-----SCALLDSLT-----GV 6080
 Db 5390 NAOFTCTIPGVKPTISWYKGAIS-----NGAGITCTQKTSLPKYDYVFE 5438

[illegible]

RECEIVED

REF ID: A38096

71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-3634
 A:Cross-references: EMBL:X62515
 R:Tryggvason, K.
 submitted to the EMBL Data Library, October 1991
 A:Reference number: S77946
 A:Accession: S77946
 A:Molecule type: mRNA
 A:Residues: 1-57, 'D', 59-434, 'A', 436, 'E', 438-449, 'Q', 451-502, 'A', 503-792, 'K', 794-908, 'R', 910-911, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 3633-3634
 A:Cross-references: EMBL:X62515; NID:G29469; PID:CAA44373.1; PID:G29470
 R:Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
 Genomics 11, 389-396, 1991
 A:Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the cDNA
 A:Reference number: A41059; MUID:92120660; PMID:1685141
 A:Accession: A41059
 A:Molecule type: mRNA
 A:Residues: 'RT', 892-908, 'R', 910-1101, 'L', 1103-1132, 'L', 1134-1221, 'L', 1223-1397, 'KX2>
 A:Cross-references: GB:S76436; NID:G243370; PID:AA21121.1; PID:G243371
 R:Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo, R.V.
 Genomics 10, 673-680, 1991
 A:Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellular localization, and cDNA sequence
 A:Reference number: A40306; MUID:91365376; PMID:1679749
 A:Accession: A40306
 A:Molecule type: mRNA
 A:Residues: 1018-1405, 'G', 1407-1409, 'G', 1411-1465, 'DOD>
 A:Cross-references: GB:M54283; NID:G184424; PID:AAA52699.1; PID:G184425
 R:Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den Berghe, H.
 J. Cell Biol. 109, 3199-3211, 1989
 A:Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal antibodies
 A:Reference number: A33625; MUID:90078352; PMID:2687294
 A:Accession: B33625
 A:Molecule type: protein
 A:Residues: 1379-1384, 'X', 1386-1388, 'X', 1390-1398, 'HE2>
 A:Accession: A33625
 A:Molecule type: protein
 A:Residues: 2166-2171, 'X', 2173-2175, 'X', 2177-2185, 'HE3>
 A:Note: peptide potentially matches four different regions of sequence shown
 C:Genetics:
 A:Gene: GDB:HSP62
 A:Cross-references: GDB:126372; OMIM:142451
 A:Map position: 1p36.1-1p36.1
 C:Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repeat homology
 C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembrane protein
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-4391/Product: perlecan #status predicted <MAT>
 F:22-193/Domain: I <DOM1>
 F:194-530/Domain: II <DOM2>
 F:198-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
 F:531-1676/Domain: III <DOM3>
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG>
 F:1563-1610/Domain: laminin-type EGF-like homology <EG7>
 F:1613-1668/Domain: laminin-type EGF-like homology <LEG8>
 F:1677-3686/Domain: IV <DOM4>
 F:2007-2034/Domain: transmembrane #status predicted <TRM>
 F:3687-4391/Domain: V <DOM5>
 F:3845-3880/Domain: EGF homology <EGF1>
 F:3888-3921/Domain: EGF homology <EGF>
 F:3953-4106/Domain: laminin G repeat homology <LG2>
 F:4149-4151/Region: motor neuron attachment (L-R-E) motif
 F:4299-4301/Region: motor neuron attachment (L-R-E) motif
 F:65.71.76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
 F:89.554, 1755, 2121, 3072, 3105, 3279, 3780, 3836, 4069/Binding site: carbohydrate (Asn) (covalent)
 F:2995, 3933, 4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 2.6%; Score 1074; DB 2; Length 4391;
 Best Local Similarity 24.2%; Pred. No. 4e-24;
 Matches 653; Conservative 273; Mismatches 947; Indels 830; Gaps 140;

QY 32 QIVGNPTPOVSEKQOQPVVAGARPLAQ-----DQDLVRLTILDLALGDSQY 80
 DB 1436 QITGNMILVA-----SQPALQGPERRSYEIFREEFWRPDPQPATREHLLMALADDEL 1491
 QY 81 VCRARNAICEAFAAAGLVQDAEAACAEQA-PHFLLRPTSIRVREGSEATFCR-----VG 134
 DB 1492 LIRA-----TFSSVPLVASTASVLEVAQGPSNRPRALEVE-----CRCPGPIG 1538
 QY 135 GSFR---PAVMSKDGRLG-----EPDG-----PRVVRBELGEASALRIRARPRDG 180
 DB 1539 LSCQDCAPGYTGTSGLYLGHCELCNCHSGLCHP-----ETGACSCQHNAA-----GE 1589
 QY 181 TYEVRANPLGAASAAALVWDSDAADTASRPGTSTAALLAHLQRRREARAE-----APA 237
 DB 1590 FCELCAPIYGDATAGT-----PDCQPCACPLINPENNFS-----RTCESLGAAGVRCAT 1641
 QY 238 SPFSTG-----TRTCTVTEGKARLSQYV 261
 DB 1642 EPGYTGQYCEQCGPGYGNPSVQGGCLPETNQAPLAVVEVHPARSIVPQGGSHS-LRCQV 1700
 QY 262 TCEPKPETVW-KKQQLVTEG---RRHVYDAQENFVLKILFCKOSDGLYCTTASNL-V 317
 DB 1701 SGSPPHYFYWSHEDGRPVFSGTQQRH-----QGSLEHFFSVQPSDAGVYCTICRNLUH 1753
 QY 318 GQYSSVLVVRPAVPFVKRLQD---LEVREKESATFLC---EVPQPSSTEAAMFKEETRL 372
 DB 1754 SNTSRAELLVTEAPSKPITVVEEQBSQSVRPGADVTFICTAKSKSPAYTLVW-----TRL 1809
 QY 373 WASAKYIEEBETERLVRNVNSADDDAVYICETPEGSRVVAELAVQGNLRLKLPRTA- 431
 DB 1810 -HNGKLPTRAMDFFNGILTIRNVQLSDAGTYVC-----TGSNNFAMDQGTAT 1854
 QY 432 --VRVGDITAMFCVELAVPVPVHMLRNQEEVAGGRVAISAEGRTHLTITISQCCLEDVQ 489
 DB 1855 LHVQASGT-----LSAPVVSIIH-----PPQLTVQP-----GQ 1881
 QY 490 VAFMAGDCQTSRTRFCVSAAPRPPLOPPVDPVVKARMSSVILSWSPPHGSRP--VTIDG 547
 DB 1882 LA-----EPRCSATGSP-----TPTLEWTGGPGQLPAKAQIHG 1915
 QY 548 YLVEKKLGLTYTWIRCHEAEWVATPELTVADVAEENFQFRVSNALNSGQSPYLEFPCTV 607
 DB 1916 GLIR-----LPAVEPTDQA-----QYLCRAHSSAQO----- 1942
 QY 608 HLAAPLAVRTPPLKAYQAVEGG---EVTFSDVLTVASAGENFLDQALKASSVYEHCDRT 664
 DB 1943 -----VARAVLHVHGGGPRVQVSPERT-----QVHAGRT 1972
 QY 665 RHTLTIREVPASLHAQKLFVANGIESIRMEVRAAPOLGTANKPAAAAREVLARLHEEA 724
 DB 1973 -----VRLYCRAA----- 1980
 QY 725 QLLAELSDQAAAVTWLKDRTLSLPGPKYEQVQASAGR---RVLLVVDVARDADAGLYECV- 779
 DB 1981 -----GVPSATITWRKEGSLPP-----QARSDTATLIPAITADAGFYLCVAT 2028
 QY 780 SRGGRIATVQLSVOGLARFLHDKMAGSCVDVAVAGGPAQFECET-----SEAHVHHWY 831
 DB 2029 SPAGTAQARMQVVLISA---SDASPPGVKIESSPSFVTEGQTLDLNCVVASAIAQVWTWY 2085
 QY 832 KQGMEL-GHSGBERFLQEDVGTTRHLVAATVTRQDEGTYSQV---GEDSVDFRLVSV- 884
 DB 2086 REGGSLPHT-----QVHSGRLRL--POVSPADSGEYVCRVENGSGPKESITVSVLHG 2137
 QY 885 -----EPKVVFAKQLARRKQLAEGASATLSCEV-AQAQTEVTWY 924
 DB 2138 THSGPSYTPVPGSTRPIRIEIPSSSHVAB-----GOTLDLNCVVPQQAHAQVTH 2186
 QY 925 KDGKLLSSSKVCMKMEATCTRLRVVQQAQADAGEYSCERAGQRLSFHLDVYK---EPKVV 981
 DB 2187 KGGSLPARH-----QTHGSLRL--HQVTPADSGEYVCHVGTSGPLEASVLTVEASVI 2240
 QY 982 ---FAKDQVAHSEVOAEAGANATLSCEVA-QAQAEEVVMYKDGKLLSKVHVHBAKGR 1037

Db 2241 PGPIPPVRIESSSTVAEGQTLDLSCVVGQAHAQVWYKRG-----GSLPAHQVGR--S 2294
 Qy 1038 RLWVQQAQKTDAGDYSCBA-RGQRVSFLRHIT-----EPKMPFAKE 1077
 Db 2295 RLYTFQASFPADAGQVVCASNGMEASITVTGTQGANLAYPAGSTOPRIEP----- 2347
 Qy 1078 QSVHNEVQAEAGASAMLSCEV-AQAOTEVWYKDGKLLSSSSKVGMEVKGCTRLVLPLQA 1136
 Db 2348 -----SSSQVAEGQTLDLNCVVPQSHQAQVTHKRG-----GSLPVRHQTHGSLRLY--QA 2397
 Qy 1137 GKADAGEYSCAGGQRVSFLHII---TEPK-----GVFAKEQSVHNEVQAEAGTAMLS 1187
 Db 2398 SPADSGEYVCRVLGSSVPLEASVLVITIEPAGSVPALGVTPTVRIESSSSQVAEGQTLDLN 2457
 Qy 1188 CEVA-OPCTEVTWYKDGKLLSSSSKVRMEVKGCTRLVLWQVQKADAGEYSC-----EAGG 1242
 Db 2458 CLVACQAHAQVTHKRGSLFA-----RHQVHGSRLRL--QVTPADSGEYVCRVVGSSGT 2511
 Qy 1243 QRVSPQLHITEPKAFVFAKEQVHNEVTEA-----GASATLSCEVA-QAOTEVWYKDG 1294
 Db 2512 QEASVLVTI-QORLSGSHSQGVAVPVRIESSSSASLANGHTLDLNCVASQAPHTITWYKR 2570
 Qy 1295 GKLLSSSSKVRIEBAAGCNRQLVWQAGQADAGEYTC-----AGQRLSFHLV-----SE 1345
 Db 2571 GGSGLPS-----RHQIVG--SRLRIPOVTPADSGEYVCHVSNAGAGSRETSLIVTIQSGSSH 2624
 Qy 1346 PKAVFAKEQLAHRVQAEAGAIATLSCEVA-QAOTEVWYKDGKLLSSSSKVRMEVAGCT 1404
 Db 2625 VPSVSPPIRIESSSPTVEGQTLDLNCVVARQPAIITWYKRGSLPS-----RHQTHGSH 2680
 Qy 1405 RLVVQQAQADTGEYSCAGGQRLSFLDVABPKVFA-----KEQPVHREV 1452
 Db 2681 LRL--HQMSVADSGEYVCRANN-----NIDALEASIVISVSPSAGSPASFGSSMPIRIES 2733
 Qy 1453 QAQ---AGASTTSLCEV-AQAOTEVWYKDGKLLSSSSKVRMEVAGCTRLVLWQAGQAD 1508
 Db 2734 SSSHVAEGETDLNCVVPQQAHAQVTHKRGSLPSHQTR-----GSLRLHLHVS PAD 2787
 Qy 1509 AGEYSC-----BAGSORLSFHLHVAEP-KAVFAKEQPASREYQAEAGT 1550
 Db 2788 SGEXVCRVWGSSGLEASVLVTEASG---SSAVHVPAPGAPPIRIEPPSSRV-AB-GQ 2842
 Qy 1551 SATLSCEV-AQAOTEVWYKDGKLLSSSSKVRMEVAGCTRLVLWQAGQADAGEYSC-- 1607
 Db 2843 TLDLKCVPQQAHAQVTHKRGSLPARHQVH-----GPLRL--NQVSPADSGEYSCQVT 2896
 Qy 1608 --AGDORLSFHLV--ABPKVVFAP--EQPAHREVQAE--AGASATLSCEV-AQAOTEV 1657
 Db 2897 GSSGTLASVLVTEIPSSPGPIAPGLAQPIYIEASSSHVTEGQTLDLNCVVPQQAHAQV 2956
 Qy 1658 TWYKDGKLLSSSSKVRMEVAGCTRLVLWQAGQADAGEYSCAGG-----QRLSFLHVA 1712
 Db 2957 TWYKRGSLPA-----RHQTHGSLRL--HLVSPADSGEYVCRASGFGPEQASFTVTP 3010
 Qy 1713 ELE-----POISERPCREPLVWKEHEDIILTATLTPSA-ATVTM-----LKDG 1757
 Db 3011 PSEGSYLRSPVISIDP--PSSVTQGGQDASFCLIHGAAPISLEWTRNQIELEDNV 3067
 Qy 1758 EI-----RRKRHET-----ASQ-----DTHTLVHGQAVLDLSAIYSCRVAEG 1797
 Db 3068 HISPNGSITITVTRPSNHGTYRCVNASNAYGVAQSVNLSVHGPTVS-----VLPEG 3120
 Qy 1798 QDFPQVEEVAAKFCLLEPVCGLGGTGTTLACELSPACAEVWVRGNTOPRVKGRFQMV 1857
 Db 3121 -----PWWKVGKAVTLEC-----VSAGEPRSSARWRI 3149
 Qy 1858 AEGPVR-----SLTVGLRAEDAGEYVCSRDDHTSAQLTVSV-----P 1896
 Db 3150 SSTPAKLEQRTYGLMDSHAVLIQISSAKFSDAGTVCLAQNALGTAQVQVEIVDTGAWAP 3209
 Qy 1897 RVVFMGSLSVVAAEGGEATFCQVSPSPDVAVV-WFRDQALLOPSKFAISQSGASHSL 1955

Db 3210 GAPOVQAEBAELTVEAGHTATLRCSATGSPAPTIHWSKLSPLPWQHRL-----EGD 3263
 Qy 1956 TISDLVLEADAGQITVBAEGASSAALVRREAPVLFKKLEP--QTVVEERSSV---TLEV 2009
 Db 3264 IIPRAQDSGGQYICNATSPAGHA-----EATILHVESPPYATTVPHEASVQAGETVOL 3318
 Qy 2010 E-LTREWPEL--RWTRNATAL---APGKNVEIHAEGARHRLVLHNVGFPADRGFFGCETPD 2063
 Db 3319 QCLAHGTPTLTQWSRVGSSLPGORATARNELHFERAAPE-----DSGRYRCRVTN 3369
 Qy 2064 DKTOAKLTIVEMRQVRLVRG-----LQAVEAREOGTATMEV--QLSHADVDS 2108
 Db 3370 KUGSABAFQAL-----LVQGPGLPATSPAGSTPTVQTPQLETKSIASVEFFHCAVPS 3425
 Qy 2109 -----WTRDGLRFOQGPTCHLVRGPMHTLTSLGRPEDSGLVVFKAEG---VHTSA 2157
 Db 3426 DQOTQLRWFKEG---GQLPPGHSVQDG---VLRIQNLDOSCOGTTCQAHGFWGKAQASA 3479
 Qy 2158 RLVVTELP---VSFSRPLQDVVTTTEKEKVTLEC-ELSRPNVDVRMLKDGVELRAGKTMAI 2213
 Db 3480 QLVIALPQSVLNIIRTSVQTVV--GHAVEFECLALGDPKQVTKSVKVGHLRPG--IV 3534
 Qy 2214 AAGACRSITIVRCFADQGVVCDHDAQSSASVKVQGRITLYIRRVLASDAGEIQV 2273
 Db 3535 QSGGVVR---IAHVELADAGQYRCATNA----- 3560
 Qy 2274 AENAESRAQLRVKELPVTLVRLDRDKIAMKHEKRGVLECOVSRASA-----QV 2320
 Db 3561 AGTTQSHVLLVQALP-----QISMPQ-----EVEVPAGSAVFPICIASGYTPDI 3606
 Qy 2321 RHFKSGQELQOPKPYELVSDGLYRKLITSDVHAEDDTYTCD---GDVKTSAOFFVEE 2376
 Db 3607 SWSKLDGSLPPDSRLR-----NNMLLPSPVPQDAGTYVCTATNRQGVKAFALQVPE 3660
 Qy 2377 QSI 2379
 Db 3661 RVV 3663

RESULT 12

S20901

titin - rabbit (fragment)

C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999

C;Accession: S20901; I46520

R;Label: S.; Gautel, M.; Lakey, A.; Trinick, J.

EMBO J. 11, 1711-1716, 1992

A;Title: Towards a molecular understanding of titin.

A;Reference number: S20897; MUID:92258380; PMID:1582406

A;Accession: S20901

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 1-6805 <LAB>

A;Cross-references: EMBL:X64696

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992

R;Label: S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Hsieh, C.L.; Francke, U.;

Nature 345, 273-276, 1990

A;Title: A regular pattern of two types of 100-residue motif in the sequence of titin.

A;Reference number: I46520; MUID:90238553; PMID:2129545

A;Accession: I46520

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 4235-5250 <LA2>

A;Cross-references: EMBL:X17329; NID:gl756; PIDN:CAA35207.1; PID:g930251

C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pr

C;Keywords: muscle

Query Match

Best Local Similarity 2.5%; Score 1046.5; DB 2; Length 6805;

Matches 1394; Conservative 968; Mismatches 2748; Indels 2411; Gaps 346;

Qy 17 KAFVSVGKDATLSCQIVGNPTTPQVSW-----EKDQOPVTAGAFRLAQDGLRL 67

Db 476 KLVVVRAGCIRLPAIVRGRPAKVTRKVGIDNVRKQGVLDVDMFAP-----L 525
Qy 68 TILDALGDSGOVYCRARNAIGFAAAGLOV-----DAEAAAC----- 105
Db 526 VIPNSTRDSDGKYSLTILVNPAGEKAVFVNRVLDTGCPVSDLKVDVTKTSCHVSWAPPE 585
Qy 106 ---AEQAPHEL-----LRPTSIRYRE---GSEATERCRV----- 133
Db 586 NDGGSQVTHYIVKRDABERTWSIVNPEVKTKSCQTNLVPGNEYFRVTVAVNVEYGGVP 645
Qy 134 GGSPPRAVSKGQRRRLCEPDGPR--VRVEELGEASALRIRAAFRPDGGT-----YEVRA 186
Db 646 ADVKPKVL--ASDG--LSEPDPPKLEVTWTKNSATLAWPLRDLGGAKIDGYIISYRE 701
Qy 187 ENPIGAASAAAALVVDSDAADTASRPG-----TSTAALLAHLQRRREARAGAPASP 239
Db 702 EDQPADRWTEYSVVVDLSLVLITGLKEGKYKFRVAARNAVGVSLPREAGVBAKEQLIP 761
Qy 240 PS-TGTRCTVTEGKHARLSYVTGEPKPTWKK-DGOLVTEGRHRVVEDAQENFLK 297
Db 762 PKILMPQITIKAGKLRIRBAHVVGKPOPICKWKGEDDVVTSS--HLAVHKAESSILI 819
Qy 298 ILFCQSDRGLYTCTANLVGQTVSSVLVVRE---PAVPFKRLQDL----- 342
Db 820 IKDVTBKSDGYSLTAENSSGTDQIKIVMDRPGPPQPPFD--ISDIDADACSLSMHI 877
Qy 343 -----EVREKESATFLCEVPO-----PSTEAAW-FKEETR----- 371
Db 878 PLEDGGNIYIIVEKCDVSGDWVTALASVTIKTSRICKLIGQEVFVRVRAENRFGIS 937
Qy 372 -----LWASAKYIGIEEGTERRLTVRNVSAODDVAIYC-----ETPEGSRTVAELAVQG 420
Db 938 EPLQSPKMLAQFPGVPSEPKNARVT---KVNKDCIFVAMDPRDSDGGS-----PITG 987
Qy 421 NLLRKLPRKTA--VRVGDTFMFCVELAVPGVPHLNRQEVNAGGVAISAGTRHTLT 478
Db 988 YLIERKGRNLLWYKANDTA-----VRSTEYPGAG-----LVGGLYSFR 1027
Qy 479 ISQCLEDVGQVAFMAGDCQSTRFVCSAPRKP-----PLQPPVDPVVKARMESSVI 530
Db 1028 I--YALNAGS-----SPSPKPEYVTAETPDPGKPEVIDVTKSTVS 1069
Qy 531 LSWSPPPH--GERPVITDGLVLEKKLGTVTWIRCHEAEWATP-----ELTVADVAEE 582
Db 1070 LIAWAPKHGDSK---IIGYFVEACKLPGDKWVRN-----TTPHOIPEBEYVTGLEEN 1121
Qy 583 GNFOFRYSALN--SFGOSPYLERPGTVH--LAPKLAVRTPKAVQAVEGGEVTFSDVLT 637
Db 1122 AOYQFRAIKTAVNISQPSSELRTPVTTHAENVPRIDLVSAMKSLLTVKAG--TNVCLDAT 1180
Qy 638 V-----ASAGEWFLDQALKASSVYIHCDRTRHTLTIREVPASLHGAKLFVANGIES-- 691
Db 1181 VFGKPMPTVSWKSGTVLKPAGGIKMAQNRNLCTLELFSVNRKDSGDYTTIATENSNGSKS 1240
Qy 692 -SIRMEVRAAPGLTANKPPAAAAREVLARLHEEAQALLA---ELSDQAAVT--MLKDGRTL 746
Db 1241 ATIKLVLDRPG-----PPASVK---INKMYSDRMLSWPPELDDGGSEITNIVDKRET 1292
Qy 747 SPGPKEYEQA-----SAGRRVLL-----VRDVARDDAGLYCEVSGRGIAYQLSVOGLA 795
Db 1293 SSRNMAQVSANVPITSCSVEKLIBGHEYQFRAICENKYGVGDPVFTBPAIAKN----- 1345
Qy 796 RFLKHOMAGSCVDVAGGPAQFECETSEAHVHVHVKYKDGWELHSGERFLQEDVGRHRL 855
Db 1346 ---PYDPPGRCDPPVIS-----NVTXDMHTVSWKPPADDDGSGPITGYLLEKRET--HAV 1394
Qy 856 VAAVTTRQD--EGTYSRCRVGSDSVDFRLVS-----EPKVVFAKEQJ----- 895
Db 1395 NWTKVRNKPVIERTIKATGLQEGTEYFRVTAINKAGPKPSDASKAVYAQDPLPPGPP 1454
Qy 896 ARRKLOABAGASATLS-----CEVAQAOTEVWYKDG--KKLSSSKVC 937
Db 1455 AFPKYDTRTSVLSNGKPAYDGGSPPIIGYLVVEVKGADTD--NWRNCLNPKLU---OKTR 1510

Qy 938 MEATGCT-----RRLVVQAGQADAGEYSCEAGGQRLSFHLVDVKEPKVVFPAKDOVAHS 990
Db 1511 FEVTGLMENTEYQFRVYAVNVKGYSDPSDVP-----DKHCPKOILLIPPE---G 1555
Qy 991 EVQAE-----AGANATLSCEV-AQAQAEVMMYKDGKLSLSSKLVHVEAKGCRRLVV 1041
Db 1556 ELDADLRKTLILRAGVTWRLVVPVKGRPPPKITWSKENVNRLRERIGLDIKSTDFTDLRC 1615
Qy 1042 QOAKTDAGDY-----SCEARGQVSFRHLIT--EPQMFPAKEQS----- 1079
Db 1616 ENVAKYDAGKYLITLSENSCGKGYTIIVKVLDTGPPVNVTVKEISRDSAYITWDPPIVD 1675
Qy 1080 -----VHNEVOAEAGASAMLSCEVAQAOTEVWYKDGK-----KLSSSKVMGMEVKGC 1127
Db 1676 GSGPIINVVKRDAERKSNSTVTECKTSFRVSNLEBGSYFFRVYAENEYIGIDPGE 1735
Qy 1128 TRRLVLPQAGKAD-----AGEYSCEA-----GQQRVSFHL--HITEPKG 1164
Db 1736 TR-----DAVKASETPGPVVDLKVLTVTKSSCNIGWKKPRSDGSRITGYVDFLTE-- 1787
Qy 1165 VFAKEQS VHNEVQAEAGTAMLSCEVAQPOTEVWYKDGKLS--SSSKVRMEVKGCTRR 1222
Db 1788 -----ENKQVRMKSLSLOYSTKDLNEGKQVTFR 1816
Qy 1223 LVVQ-QVGKADAGEYSCEAGQQRVSFQHLITE-PKAVF-AKE-----QLVHNEVTRTEAG 1273
Db 1817 VSAENENGEGTPEITVVAKDDVAPDLDLKDLPCYLAKENSFRLLKIPHQ----- 1869
Qy 1274 ASATLSCEVAQAOTEVWYKDGKLSLSSSKVRIEAGCMQLVVOQAGQADAGEYTCEAG 1333
Db 1870 -----GKPAFSPVTKWKGEDPLATIDTRVSVSSAVNTVTVYDCCKDAGKAIT-- 1916
Qy 1334 GQRLSFHLDSBPXAVFAKEQLAHRKQVABAGATLISCEVAQAOTEVWYK-----DG 1387
Db 1917 -----ITLKNVAGTKEGTSLIKVGKCPG-IPTGPIKFEVTAETLTKWGPXKDDG 1966
Qy 1388 KK-----LSSSKVRMEAVGCTRLVVQACQADTGEYSCEAGGQRLSFLDVAEPKV 1441
Db 1967 GSEITNYILEKSDSVNNKWVTC-----SAVQKTFTRVRLHEGMEYTFRVS-AENKYG 2019
Qy 1442 FA---KEOPV-----HREVOQ-----AQAGASTLISCEVA-QAQ 1470
Db 2020 VEGGLKSEPIVAKHPFVDPDAPPNPINVDVRHDSVSLTWDPRTKGTGSPITGYHIEPKER 2079
Qy 1471 TEVMMYKDGKLSFSSKVRMEAVGCTRLVVQAGQADAGEYSCEAGSQRSLSFHLHVAEP 1530
Db 2080 NSLLWKRANK-----TPIRMKDFKVT-----GLTEGLEVEFRVMAINLA--GVGKP 2123
Qy 1531 K-----AVFAKEQBPASREYQAEAGTSATLSCEVAQAOTEVW-----YKDGKLS-- 1576
Db 2124 SLPSFVVALDPIPPGKPEVINTRNSVTL-----INTPEKYDGHKLTGYIV 2172
Qy 1577 -----SSKVRMEA-----VGCTRLVVQAGQADAGEYSCKAGDQRLSFLHVAEPKVVF 1626
Db 2173 EKRDLPSTKMTWKANHNVPDCA--FTVTDLVEGGKYEFRIKAKNTAGATSAISESTGTII 2230
Qy 1627 AKEQ-----PAHRE-VQAEAGASATLSC--EVAQAOTEVWYKDGKLSLSSSKVRV 1674
Db 2231 CKDEYEAPTIVLDPFTIKDGLTIRKAGTIVLNAISILGKPLPKSSWSKAGKDIRSDITQI 2290
Qy 1675 EAVGCTRLVVQAGQADAGEYSCEA-----GQRLSFLHVAELEPQISERPCREPLW 1730
Db 2291 TSTPTSSMLTVKYSRCKDAGEYITATNPFGTKBEHRVAVTVLDV-----PGPPGPIEI 2343
Qy 1731 KEHEDIILTATLATPSAATVW-----LKDGVEIRS-----KRHETAS-----QGDTH 1774
Db 2344 SN-----VSAEKAALLTTPPLEDGGSPISKYVLEKKEGTSRLLWTVVVAEDIQCRH 2393
Qy 1775 TLTVHGAQVLDIAIYSCRVGA---EGQDFPVQVEEVAAKFCRLLEP-----VCGELG 1823
Db 2394 VT-----KLIQNEVLFVSAVNHVKGEPVQSEPV--KMVDREPGPPGPKPEVSNVTK 2447

| | | | | | | | |
|----|------|--------------|--|--------------------------------|------------------------------|-------|------|
| Qy | 1824 | GTVTLACE | -----LSPACAEVVRWRCNGTQPRVKRFQWVAEGVRSI | -----TVJGLR | ----- | 1871 | |
| Db | 2448 | NTATVSWKRPTD | GGSEITGYVYERREKKGLRWVRAITKTFVSDLRCKVTQ | LEGTNYEYFR | 2507 | | |
| Qy | 1872 | ----- | AEDAGSY | -----VCESRDDHTSAQLTVSVPRVKFMSG | LTVVVAEBGGE | 1915 | |
| Db | 2508 | VSAENRAGIGP | PSDASNYLWKOVAYAGPPSNARVDTTKKSASLAWGPHY | -----DGLL | 2564 | | |
| Qy | 1916 | ATFOCVVPSDVA | -VWFRD-----GAL----- | LOPSEKF----- | AISQSGASHLTI | 1957 | |
| Db | 2565 | EITGYVVEHOK | VGDETWMKDTTGPALRITBFBVFDLTKKYNFRISAIN | DAGVGEPAVI | 2624 | | |
| Qy | 1958 | SDVLVEDAGOI | TVAEAGASS--SAAURVREAPVLFKKLEPQVBERSS | SVTLEVELT-- | 2012 | | |
| Db | 2625 | PDVEI----- | VEREMAPDFELDAEUR | ----- | RLTVVRAGLSIRIFVPIK | 2663 | |
| Qy | 2013 | RPWPELRTRN | ATALAPGKNVE----- | IHAEGARH----- | LVLHNVGFADRGFFG | 2058 | |
| Db | 2664 | GRPAPEVTWK | DINLTRANIENTESFTLLIIP | ECNRYDTGKFVMTIENPAGKKSQFVN | 2723 | | |
| Qy | 2059 | CET----- | PDDKTQAKLTVENRQVRLVRG | ----- | LOAVEAREOGTATMEVQ | 2099 | |
| Db | 2724 | VRVLDTPGV | VLNURPTDITKDSVTLHW-DLPLIDGGSRITNYI | VEKEATRKYSVSTVTKK | 2782 | | |
| Qy | 2100 | LSHADVDG | SWTRDGLRFQOQPTCHLAVRGPWHITL | SLGRPEDSGLMVFVKAEGVHTSARL | 2159 | | |
| Db | 2783 | ----- | ----- | CHKC----- | TYKVTGLSGCEYFFRVVVAENEYGI | 2814 | |
| Qy | 2160 | VVTELPVSR | PLQDVVTTKEKVTLECELSRPNVDVRWLK | ----- | DGVELRAGKTMATAAQ | 2216 | |
| Db | 2815 | SETKEPVKASE | -----APSPDLSNI-MDITKSTVSLAWPKPHD | GSKITGVYIEAQRK | 2868 | | |
| Qy | 2217 | GACKSLTIY | RCEFAQGVVYCDADHADQASSAKVQGR | TYTLIYRVRVLAEDAGBIQFVAEN | 2276 | | |
| Db | 2869 | G----- | SDQWTHITTVKGLBCVVRNLTEGEYT | ----- | FQWMAVN | 2903 | |
| Qy | 2277 | AESRAQUR | -----VKE---LPVTLVRPLRDKTAMEK | HRGVLECOV--- | SRASAQVRWFXX | 2324 | |
| Db | 2904 | SAGRSAPRES | RPVTKBQTMPELDRGIYQKLVIAKAGDN | IKVEI | PVLGRPKPTVTWKX | 2963 | |
| Qy | 2325 | GSQELQGP | KVELVSDGLYRKLIISDVHAEDEDTYT | CDAGDVKTSAFFVEEQSITIVRG | 2384 | | |
| Db | 2964 | GDQVLKQTR | VNVENTATSTILNISECVRSDSGPYPL | TAKNI----- | VCEVGDVJTIQ | 3016 | |
| Qy | 2385 | LQDVTVME | PAPAFECETSPSV----- | RPPKLLGLTKTVLQAGNVGLEQGV | ----- | HLR | 2436 |
| Db | 3017 | VHDIPGPTG | PIKFD-EVSDGFVTFSEPP----- | ----- | ENDGGVPISNYVI | 3058 | |
| Qy | 2437 | MLRITCSMT | GPVHFTVGKSRSSARLVSDIPVVL | TRPLBPKTRHQLQSVLVSCDPRPAP | 2496 | | |
| Db | 3059 | EMRQTDST | ----- | ----- | ----- | ----- | 3082 |
| Qy | 2497 | KAVQWYKDD | PTLSPSEKFKMSLEGOMAEILRLMPAD | AGVYRCQAGSAHSSTEVTVEAR | 2556 | | |
| Db | 3083 | ----- | IRLITTVGEYQFRVKAQ | ----- | NRVYGVGPGITSASI-VANY | 3116 | |
| Qy | 2557 | EVTVTGLO | QAEATEBGWASFCSELSHEDREVEW | ----- | SUNGHMPLYNDSFHEISHKGRRH | 2612 | |
| Db | 3117 | PFKVP | GGPTQVPT----- | AVTKDSMTISWHEPLSDGGSPILG | -----YHVBERK-ERN | 3163 | |
| Qy | 2613 | TLVLKSTORA | -DAGIVRASLSKYSTEARLEVR- | ----- | KPVVFLKALDLSA | 2658 | |
| Db | 3164 | GILMOTVSK | ALVPGNI | FKSSGL--TDGIAVEFRVIAENMAGKSKPS | RSPSPVIALDFIDP | 3221 | |
| Qy | 2659 | EEBRTLAL | CEQVSDPEAHVVRWKGQVQLGSPDKYDFL | HTATGRGLVHVHDSVPEDAGLYTC | 2718 | | |
| Db | 3222 | PGK--- | BIPLNITRHTVTLKWAK | ----- | PEYTG--- | 3246 | |
| Qy | 2719 | HVGSSEET | RRVVRVHDLHVGITKRLKLTWEVLEGESCS | FECVLSHESASDP | PAMWTVGGKTVG | 2778 | |
| Db | 3247 | --GFKITSYI | VEKKDJPNRGLKANFN | NIENE----- | FTVSGUTED | 3286 | |
| Qy | 2779 | SSSRFOAT | QGRKYILVVR | EA----- | PSDAGEVYVVRGJLTKSASLI-VRERPA | AIKIP | 2832 |

[illegible]

Db 4131 GSSINNYVEKRTSTTTWQIVSATVARTTIKASRLKTG-----CBYQFR-IA 4178
 Qy 3776 VVDAGEYSCVGOBERTSATLTVR-----ALPARFIEDVK-NQAREGATAVL--Q 3822
 Db 4179 AENYRGKSTVYNSBPVIAQYPFKVPVGGPTGFVFTLSSRDSMEVQWNEPVNDGGRVIGYH 4238
 Qy 3823 CELSKAAPVWRKSEITLGGDRYSLR--QDGTCELOIHLGSLVADTGEYSVCVGOBERTS 3880
 Db 4239 LERKERNILWVKNLKTPIQTKFTTGLBEGIEYEFVSAENIVGIGKPS-----KVS 4292
 Qy 3881 ATLVTRAPQPVFREPLQSLQAEAGSTATLOCELSSEPTATVWMSKGLQIANGRRRPRLQ 3940
 Db 4293 ECVYAROP-----CDPPRPERIIVTNSVTLQ---WKXPTVD 4327
 Qy 3941 GCTAELVLQLOREDTGEYTCGSOQATSATLVTAAPVFLRELQHQVDEBGTALCC 4000
 Db 4328 G-----GSKITGYV-----VEKKELPDG-----4345
 Qy 4001 ELSRAGASVWRKGSLOLFPCKAYO---MVQDGAABELLVRGVEQEDAGDYTCDTGHTOS 4057
 Db 4346 -----RWKASFTNIMDTQEVTLVED---HRYEFVRIARNAAGVSEPSSESTGA 4393
 Qy 4058 M-ASLSVRVRP-----PKFKTRLQSLQETGDIARLCCQLSDABSGA-----VWOMLKEGVE 4108
 Db 4394 ITARDEIDPPRISMDPKYK---DTIVVHAGESPRI-----DADIYKPIPTTQWIKGDQ8 4445
 Qy 4109 LHAGPKYEMBSOGATRELLIHOLEAKDTGEYAC---VTGGQKTAASLRVTE---PEVT 4160
 Db 4446 LSNTRALEIKSTDPAFSLSVKDAFRVDSGNVYLKQAVAGERSVTVNVKLDVRGPPPEP 4505
 Qy 4161 IVRGLVDAEVTADDEVFCEVSRAGATGVWCQLQGLPLOSNEVTEVAVRDGRIHTLRK 4220
 Db 4506 IV-----IS 4509
 Qy 4221 GVTPEADAGTVSFHLGNHASSAQLTVRAPEVTILEPQDVQLSBOQDASFCRLSRASGOE 4280
 Db 4510 GVTAEKC-TLAWK-----PPLQD-----4526
 Qy 4281 ARWALGGVPLQANEMDITVEQGTLLHLTLKHTLEDAGTVSPHVGTCSSEALQKVTAKN 4340
 Db 4527 -----GG---SDIINYIVERRETSRLVW---TVVDA-----NVQTL8---CKYT---4561
 Qy 4341 TVVRGLNVALEGEALFECQL-----SQPEVAHWTLLDDP-----VRTSE 4384
 Db 4562 -----KLLEGNEYIFRMAVKNYGVGEPLSEPEVIAKNPFPVDPAPKAPECTVTVK 4612
 Qy 4385 NAEVVFENGLRHLLLNLRPODSQCVRTFLAGDMVTSAPLTVRGNRLEILEPLKNAVR 4444
 Db 4613 DSMIVWE-----RP-----ASDGGSEILGY---VLEKXDEGIR 4644
 Qy 4445 AGAQAFTCTLSEAVPVGEASWYINGAAVOPDSDMTVTADGSHQALLRSAPPHHAGEV 4504
 Db 4645 -----WTRCHKRLIGELRLRVTG-LIENHNVEFRVSAENA-----AGLSEPPSPSAY 4690
 Qy 4505 TFACRDVASARLTVLGLPDPPEAEVVAHSHVTLNSAAPSMDGSGGICGVYRVVEK8 4564
 Db 4691 QKAC-DPIYK-----PGPPNNPKVMDITRSVFLWSKPIYDGGCEIQGIVBEKDV 4741
 Qy 4565 ATGQWRILCHEL--VPGPECVVDGLAPGETYFRVAAVGVPGAGEPVHLPQTVALAEPPXP 4622
 Db 4742 SVGEWMTCTPTGINKNTNIEVKLEKHEYNFRICAVNKAAGVGDHADVPGPVIVE8---4797
 Qy 4623 VPPQPSAPESQVAAGEDV8LELE-----VVA8AG-----FVIWHKGMERI 4663
 Db 4798 ---KLEAP-----DIIDLUELKRIINIRAGGSLRLFVPIKGRPTPEVRKGVGDGEI 4845
 Qy 4664 QPGGRFVWVSGRQOMLVIKGF8AEDQGEYHCHGLAOGSICPA8-ATFOVALSPASVDEAP 4722
 Db 4846 RDAAIIDST8SFT8--LVLNPNVNRD8GKYTLTLENSSG8TKSAFVTVRVLDT8---SP 4898
 Qy 4723 QPSLPP8A8EGDLHLWEA-----LARKK8SR8P-----TLDS 4757
 Db 4899 PVNLKVTEITKDSV8ITWE8P8LLDGG8KIKNYI8VEK8ED8TRK8YAA8VTVN8CHK8SWKIDQ 4958

Qy 4758 ISE-----LPED8GR8QRLP-----QB8EEV8PDL8EGY8T8A8BL8ART8GD8AD8L8HT 4803
 Db 4959 LOEG8SY8FRV8TA8NEY8IGI8PART8D8PIK8VA8E8VP8QP-----PGKIT8VD8D8VTR8-NSV8LS8WT 5014
 Qy 4804 SSD8ESR8AGT8F8L8V8TL-----KK8GR8PT8S8PL8K8G8AP8PS8VK8P8Q8Q8E8PL8A8VR8P8PL8D 4861
 Db 5015 KPE8D-----GG8SKI8QY8IV8E8Q8K8H8E8K8W8E8C8AR8V8K8LE8AV8IT8NL8T8Q8EE8YL8FRV-----5065
 Qy 4862 L8TK8DL8GP8S8MD8KA8V8IQ8AA8F8K8Y8K8VR8K8EQ8G8PM8F8SH8T8FG8DT8AQ8VD8AL8R8L8E8CV 4921
 Db 5066 V8V8NE8K8GR8SD8R8SL8AP8IVA--K8LV8T8EP8DV8K-----P8AF8S-----SY8SV8Q8V8Q8DL8KIE8VPI 5115
 Qy 4922 ASK8AD8VR8AR8WL8K8D8G8V8EL8TD8GR8HH8ID8Q8G8T8C8LL8I8AG8LD8RAD8AG8CY8T8Q8V8N8K8F8Q8VT 4981
 Db 5116 S8GR8PK8PT8IT8W8TK8O8GL8PL8K8Q8T8TR8IN8V8AD8SL8DT--T8LSIK8ETH8K8D8S8GH8Y8GIT8VAN8V8G8Q8T 5174
 Qy 4982 HS8AC8VW-----V8SG8SE8E8A8E8S-----SS8G8EL-----DD 5005
 Db 5175 AS8IE8IT8LD8K8PD8PP8K8PV8K8F8E8V8S8A8E8S8IT8LS8W8N8P8LV8T8G8C8Q8IT8NY8V8H8K8R8D8T8TT8V8DW 5234
 Qy 5006 AF8R8A8RL8HL8R8L8F8RT8K8SP8E-----V8S8DE8EL8F-----LS8ADE8P-----AP8EEP-----5045
 Db 5235 VS8AT8V8ART8T8L8K8VT8K8LT8G8T8E8Y8Q8F8R8IP8EN8RY8G8S8F8AL8D8SE8P8IV8AQ8Y8PK8G8P8G8T8P8F8VT 5294
 Qy 5046 -----AD8V8Q8T8RE8DE8HF8IC8IR8E8AL8T8E8AR8Q8V-----TR8Q8E8F8AT8LG8I8G 5085
 Db 5295 AT8SK8D8MV8V8Q8HE8PIN8NG8S8PI8L8G8Y8HL8ER8K8ERN8S8IL8AT8KV8D8K8S8II8HD8T8Q8K8AL8N8E8S8IE 5354
 Qy 5086 VE8IKL-----V8QG8PR8RV8EM8C8IS8KET8P8AP8VP8PE8LP8S8LL8T8SD8A8P8VL8TEL8Q8N8O---5135
 Db 5355 YE8FR8YA8EN8IV8GV8K8AS8KN8E8CY8V8ARD-----P8CD8PP8-----T8PE8I8V8K8R8E8IT8L8Q8MT8K 5405
 Qy 5136 EV8D8G8Y8P8V8D8CV8T8G8P8M8P8SV8R8W8FK8D8--KL8LED8D8HY8M8IN8ED8Q8G8H8O8LI8T8V8V8PAD 5193
 Db 5406 P8Y8D8G8S8M8IT8Y8I8VE8K8DL8P8GR8W8K8AS8FT8V8IT8-----O8FT8V8S8GL8T8EQ 5451
 Qy 5194 MG8V8R8CLA8EN8MG8V8S8T8K8A8EL8R8V8DL8T8D8Y8T8A8D8AT8E8S8Y8S8AQ8Y8LS8R8EQ8EG8T8E8ST 5253
 Db 5452 R8Y8E8FR8V8IA8K8NA8GT8MS-----K8PS8D8T-----G8P8ITA 5478
 Qy 5254 T8DE8QL8Q8V8W--EB8LR8D8Q8V8AP8GT8L8AK8F8QL8KV8G8Y8P8AP8RL8Y8FK8D8Q8P8L8T8AS8H8I8MT8G 5311
 Db 5479 K8D8E8LP8R8IS8MD8P8R8D8T8I8V8V8N8AG8E8T8F8R8E8AD8V8H8G8K8PL8T8I8E8W8L8R8G8K8E8V8E8S8AR8C8IK8N 5338
 Qy 5312 KK8IL8HT8LE8II8S8V8T8R8ED8S8Q8V8A8Y8I8SN8AM8G8A8Y8S8AR8LL8VR8PD8E8PE8K8P8AS8DV8HE8Q8LV8PP 5371
 Db 5539 T8D8F8K8ALL8IV8K8DA8IR8D8G8Q8Y8IL8R8AS8N8V8AG8S8K8S8FP8V8N8K8V8LD8R8P8G8P8G8V8Q-----V8TG 5592
 Qy 5372 R8ML8E8FT8PK8VK8K8SS8T8F8SV8K8VB--GR8PV8T8V8W8L8R8E8E8A8ER8G8VL--W8IG8D8T8P8Q8Y8T8V8AS8 5428
 Db 5593 V8T8CE8K8T-----L8T8W8S8P8L8Q8D8G8S8D8IP--HY8V8VE8K8R8E8T8S8RL8AW8T-----V8V8ASE 5634
 Qy 5429 AQ8Q8H8SL8V8LL8D8V8G8R8O8HQ8YT8C8I8AS8N--AA8GO8AL8C8AS8I-----HV8S8GL8PK8-VE8EQ8EK8V 5479
 Db 5635 V8T8NS8L8K8IT8K8L8EG8N8E8Y8FR8IM8V8N8KY8G8V8E8P8L8E8S8AP8VL8M8KN8P8V8P8G8PK8S8L8E8V8T8N8IAK 5694
 Qy 5480 B8AL8I8ST8F8T8Q8GT8Q8A8IS8A8Q8L8E8T8A8F8AD8L8G8OR8K8E8E8P8L8A8E8AL8G8H8LS8LA8E8V8G8E8E8F8L8Q8L 5539
 Db 5695 DS8MT8V8C-----N8RP8DS8G-----G8S-----I 5712
 Qy 5540 TS8Q8ITE-----MV8SA8KIT8O8AK8LO8VP8G8D8S8D8SK--TP8S8AP8RH8GR8SR8P8SS8IQ8E 5687
 Db 5713 T8GY8I8VE8K8R8DR8SG8IR8W8IK8CN8K8R8V8T8DL8R8F8V8T8GL8T8H8E8Y8E8FR8V8SA8EN8A8G8E8F8S8P8A---5769
 Qy 5588 SS8SE8SD8G8AR8E8IF8DI8V8V8T8AD8YL-----PL8G8AB-----Q8D8AT8L8R8EQ-----5627
 Db 5770 -----T8V8Y8K8AC8D8P8V8FK8P8G8P8T8N8A8V8H8V8D8T8K8N8S8IT8L8AW8K8P8I8D8G8S8E8VL 5814
 Qy 5628 -Y8VE8LD8A8A8H8PL8R8VL8RT8K8P8T8K8S8P8S8Q8G8W8V8P8AY8LD8R8RL8K8S-----P8EW8A8E8A8E8P8G 5682
 Db 5815 G8Y8I8E8I8CK8ADE8E8Q8I8V8T8P8Q8L8K8AN8R8E--IS8K8L8E8H8Q8E8K8IR8V8C8AL8N8K8V8L8G8E8A8S8V8G 5873

QY 6511 R-ERDILAAALSHPLVTGLDQFETKRLILILELCSSEELDLRYKGV-VTEAEVKVI 6568
Db 1496 RDEISIMCULHPLKVOCDFAFEKANIVMWLEWVSGELPERIIDFELTTEREIKYM 1555
QY 6569 QOLVEGHLHSHGVLHLDIKPSNLMVHPAREDIKICDFGAQNIPTAELQFSQYSGPE 6628
Db 1556 RQISEGVEYIHKQGIHLDLKPENIMCVNKTGTSIKLIDFGLARRLESAGSLKVLFGTPE 1615
QY 6629 FVSPETIIQQPVSEASDIWANGVYSILSLTSCSPFAGESDRATILNVLEGRVSSPMAA 6668
Db 1616 FVAPEVINYBPIGYETDMSIGVICYILVSGLSFPFGDNDETILANTVTSATWDFDEAFD 1675
QY 6689 HLSDDAKDFIKATQRAPOARPSAAQCLSHPWFLKSPABEAHFINTKOLKFLARSRWQ 6748
Db 1676 EISDDAKDFISLLKDKMKSELNCTQCLQHPWLQKDTKNMEAKLSKDRMKYVARRKWQ 1735
QY 6749 RLSMSYKSLIMWSRIPELLRPDPSPLGVARHLCRDTGGSSSSSSSSSSSSSSSSSSSS 6808
Db 1736 KTGHAVRAI-----GRSSMAMISGMSGRKASGSS----- 1765
QY 6809 LPPSPVTHSPHLLHPRGFLRPSA 6830
Db 1766 -PTSPINADKVENEDAFLEEVA 1786

RESULT 14
T34416
hypothetical protein F12F3.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34416
R:Fulton, B.; Wohlmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F12F3.
A:Reference number: Z21521
A:Accession: T34416
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2783 <FUL>
A:Cross-references: EMBL:U80022; PIDN:AA25886.1; GSPDB:GN00023; CESP:F12F3.2
A:Experimental source: strain Bristol N2; clone F12F3
A:Genetics:
A:Gene: CESP:F12F3.2
A:Map position: 5
A:Introns: 45/3; 90/3; 451/3; 509/1; 2313/3; 2341/3; 2378/3; 2414/2; 2453/3; 2474/2; 252

Query Match 2.3%; Score 963.5; DB 2; Length 2783;
Best Local Similarity 20.3%; Pred. No. 4.4e-21;
Matches 667; Conservative 448; Mismatches 1225; Indels 945; Gaps 145;

QY 67 LTILDLALGDSQYVCRANAIGFAAAGLVQV-DAEACAEQAPHLRFTSIRVREGS 125
Db 14 LTINSIDSQDGEFALKIKRCCGDKYAGIQVTDPRPAAPGKPAVED-QNVDVSLRWAA 72
QY 126 EATFCRVGSP-----RPAVSWSKDGRRLGEPDGPVRVVEELGEASALRIRAA 175
Db 73 PT-----NDGSPVRNYTEMCTEKGKWT-----AEVTKQAFITFLNVL 113
QY 176 PRDGTGYVRANPIGASAAALVDSADAPTASRGTSTAAHLAHLORREARBA 235
Db 114 PGESYRFRVRADNTFG-----QSEPSD-----ESELVWVKVNSR----- 147
QY 236 PASPSTGTRCTVTTEGKHARLSCVVTGPKETVWKKDGLQVTEGR----- 282
Db 148 -----VVEEPKKEVKVKEQSDYDERVAKDSEPEYKTI 182
QY 283 -RHVYVEDAQENFVLKILCKOSDRGLYTCTASNLVGTYSVLVVRREPVPFKRLQD 341
Db 183 DIHRLPNDLQAKYIIIEELGKAGYGVYRAT-EKATGKTAAKMKVQVR-PGVKKENVIHE 240
QY 342 LEVREKESATFLCEVPOPSTEAWKKEETRLWASAKYIEEBEGTEPRLTVRNSADDDA- 400
Db 241 ISMMNQLHHEKLLNLHEA-----FDMGNMWLIEFVSGGELFKILEDLSLMESEVR 294

QY 401 -----VYICETPEG-----SRTVAEL-AVOGNLLRL-PRKTAVRVD 436
Db 295 DYMHOILLGVSHMKNQIVHLDLKPENILLKAKSNELKIIDFLGLARKLDPKSKVLFG 354
QY 437 TAMFCVEIAVPGPVHMLRNQEEVAGGRVAISA-----EGTRHTLIISQCCLEDV 487
Db 355 TPEFCAPEVNYQPVGLSDMTMWTVGIVSYLLSGLSPLFGSDDEDTLANVSASDWDFDP 414
QY 488 G--QVAEMAGD--C-----QSTRFCVSAPRKPPLQPPVDPV-VKARMESV--IL 531
Db 415 SWDDVSDLAKDFICRLIMTKDKRKNVQDALBPHMITKNQPKLDKSGVPARQKRNFLSLK 474
QY 532 SMSPP--PHGE-----RPVTIDGYLVKKKLGTYTWIRCHEAEWATBELTVDVAE 581
Db 475 RMSDOLLPIGLAKRGAIFRRLTMDGVF----- 502
QY 582 EGNQFRVSALNSFGSPVLEFPGTVHLAPLAVRPLKAVCAVEG-----GEVTFVSVDLTV 638
Db 503 ERNIAPDTDA-----AP--SVKQLEDIVANVGDLIATLSCDVDGVP 542
QY 639 ASAGEWFLDQALKASSV--YEIHCDRTRHTLTIRVPASLHGAQKLFVANGIESI---R 694
Db 543 SPKVQWYKDKKELTVPSMKYDSFYNEGLAELTVKNIVESDAGKYTCRATNDLGSINTAK 602
QY 695 MEVRA-----APGLTANKPPAAAAAREVL-----ARLH-----BEAQL----- 726
Db 603 LSYKADKKKKSKTSKSPAVIEKKKORKTSKVVIEMIDMPNFHLLQDDAKIGEPKI 662
QY 727 --LAELSDQAAAATMLKDGRTLS-PGPKYEQASAGRRVLLVRDVARDDAGLYECVSRGG 783
Db 663 LVTNTTLPETVDMVHNGEHSINDSNYLKXDKGRYELHLSVDSSTDGKXKAVG--- 719
QY 784 RIAYQLSVQGLARFLHKOMAGSCVD-----AVAGGPAQFECETSEAHV----- 826
Db 720 -----KNAFGECESEAKLTWVDPQYAPSPFGKQSDVKCESDILKLE 763
QY 827 -----HVHWYKDMELHSGERFLOEDVGT-RHRLVAATVTRODEGTYSK-----RV 872
Db 764 VNTQANPAPEINFRNESIEHSOHLRQFDGSGNYSLTIIDAYAEDESSEYKCVAKNKI 823
QY 873 GEDSDVDFLRVSE-----PKVYFAKEQLARKL-----QAEAGASATLSCEVA-QAOTEV 921
Db 824 GKAAHTVCCVRIEELLSSKSKIDGSKAPRFRMQLPTPREVPQCADTLVCSVSGTPHPNI 883
QY 922 TWYKDGKLSSSSKVCWEATG-CTRELVVQQAQADAGAYSCEA---GGORLSFHLDVKE 977
Db 884 KWTKDDKPIDMSNKQVRHENGVCVTLHII--GARDDDGGRYVCEAENIHGVAOSF--SVVE 939
QY 978 PKVYFAKDVYAHSEVQ-----AEAGANATLSCEVAQAQAEVMMYKDGKKL--SSSLKV 1028
Db 940 IKAEDVKDHRPKFLPLVNCSTCEGNMVECCVTGKPIPTITWYKDGKLIENRMLQ 999
QY 1029 HVEAKGRRRLVVQQAQKTDAGDYSCEARGQVRSFLHIT-----EPKMEFA 1075
Db 1000 YTRDKGV-SRLNLMNVMNMDGDEYTCNAVNSLKGDPHTCTVWVDMGLSKRLTTPVRSKS 1058
QY 1076 KEQS-----VHNEVQ-----AEA-----GASAMLSCEVAQAQOT-EVYTWKDGKLS 1116
Db 1059 RSRSPSVVGGIQRPPVTVTRPLADATVTEGNRELLEVEVDGFPPTIEWYHDKGLVAE 1118
QY 1117 SSYKGMVEKGTGTRRLVLPQAGKADAGEYSCEAGQGVRSFLHITTEPKGVFAKE----- 1169
Db 1119 SRTLRTYFDGRVAFLLKIYEAEHEHNGQYVC-----KVSNKLGAETRAIVVVEAPDAEH 1173
QY 1170 -----QSVHNEVQAEAGTTAMLCE-VAQPOTETVWYKDGKKL---SSSSKVRMEVK 1217
Db 1174 VTQMPFTVKLQDVLVLTAGETATFTCCSYANPAQVWHLHGKALQOIKSNYKTRL-FD 1232
QY 1218 GCTRLRVQVQKADAGEYSCEAGGQ-----RVSFQLHI--TEPKAVFAKEQLVHNEVTE 1271
Db 1233 DNTATLVENVDELCTYITAVANNQDGHVHTSAQLTISGEAKIAASLPYFIILKPK 1292

```

QY 1272 AGA--SATLSCEV---AQAQTEVTWKDGKLLSSSSKVRIEAAGOMQLVVOQAQADAG 1326
DB 1293 INVVEGATLSIQADLNGSPIPEVWMLKNSSELVESDRIQMKCDGVNYQLLVDRVGLDEG 1352
QY 1327 EYTCAGGQ--RLSFHLVSEPKAVFAKEQLAHRKQVQAE-----AGAIAT 1369
DB 1353 TYTTAENEKGIKIRONTESVTKSEVKEKEKKKKEKKKQKGGKKKPGPLRPSGASKT 1412
QY 1370 LSCEVA-----QAQTEVTWKDGKLLSSSSKVRMEAVGCTRR-----LVV 1409
DB 1413 EQVTMAFDAPSEGPADSEYERRCPDQREWVSCG---STKSLELEIKGLTENTYIFRV 1468
QY 1410 QQAQADTGEYS-----CEAGQRLSFLDVAEPKAVFAKEQPAQAEAGTATSCEV-AQ 1560
DB 1469 AGKQKQGLGESEMTSTLKTASVQAQPOFTIS--PQSKIIANRDDEF--EIAVEFSGTPTP 1525
QY 1463 SCEVAQAQTEVTWKDGKLLSSSSKVR-----RMEAVGCTRLVVOQAQAD 1508
DB 1526 S-----VKYKENLQIVPDEKIDVAITSTSSILNLSQEBNGTFNCLINELQOAS 1576
QY 1509 AGEYSCE-----AGSQRSLFHLVAEPKAVFAKEQPAQAEAGTATSCEV-AQ 1560
DB 1577 A---SCQVTIFNKPASQSTPDH---SLERNLVPFLQKALNNSAQAGQOIMLTCTRISSR 1630
QY 1561 AQTEVTWKDGKLLSSSSKVRMEAVGCTRLVVOQAQADAGYSCE---KAGDQRLSF 1615
DB 1631 SESTVAMFKDRIEISAGRYELSDKSNHKLCHAVQSDTGKYRCVWTKYKGYAEBEC 1690
QY 1616 HLHVAE--PKVFAKEQPAHREVOAEAGASATLSCEV--AQAQTEVTWKDGKLLSSSSKVR 1673
DB 1691 NVAVEDVTKFTAPSFSAATLSSTAILGNITLECKVEGSPAPEVSWTKOGERISTTRIR 1750
QY 1674 V---EAVGCTRLVVOQAQADAGYSCE---AGGQELS-----FELHVAEL 1714
DB 1751 QTQDENGNC--KLSISKAESDDMGVYVCSATSAGVDSSTSMWNIANTGTQDLSHLVIAQT 1808
QY 1715 EPOISERP--CRREPLVVKHEH---DIILTATLTPSAATVTLKDGVEIRRSKR---HET 1767
DB 1809 ADEKHEKPRFTAPPSLIEVNESQOFTLIAKAVGEPKPTVTLKDGREILRTNRYHHFV 1868
QY 1768 ASQGDTHLTVHGAQVLDLSAIYSCRVCAEQDQFPV-----QVEVAAKFC 1812
DB 1869 TGDGESH-LIAECVSVKSTSGIFSCK--AENPNGFVIAETQIVQRMKPAQLANVAPKET 1925
QY 1813 RLLEPVCGEIGGTVTLACELSPACAEVVMRCGNTPQVRGKRFQWVAEGPVRSLTVLGLRA 1872
DB 1926 IPLTDMGIVNGHPITLSCNVT-----GSPPEP-----TLEWYI 1958
QY 1873 EDAGEYV--CESRDHTSAQLTVSVPRVFMGSLSTVVAEBEGEATFQCWVSPSDVAVV 1930
DB 1959 DSGHKINLTSSTTDWTECFP-----GKVAELKS--ERVLREORG--TYQCIATNSS--- 2006
QY 1931 WFRDQALQPSKFAISQSGASHSLTISDLVLEDAQGITVEAGSSASSAALRVREAPVLV 1990
DB 2007 -----CQATTQCYLLVGLSDEPAG-----PPRP 2030
QY 1991 KKLLEPQTVBERSSVTLVELTR--PMPELWTRNATAPAGKNVEIHAEGARHLVHNV 2049
DB 2031 VKCLQDWTPLKESIEFSVLAGFPPTPDLTWYHNEKKINEGKDKVI----- 2076
QY 2050 GFADRGFFGCTPDDKTOAKLTVMERQVRLVRGLQVAEAREOGTATMEVOLSHADVDSW 2109
DB 2077 -----TFPSDTS---VLSIKVSL-----ASL----- 2096
QY 2110 TRQGLRQOQPTCHLAVRGPMTLTLGLRPEDSPGLMVPKAEGVHTSARLVLTLPVSFS 2169
DB 2097 ---GMYPVEASNIHGVL-----TAGRLNVSDEKKAEPQPKHVLPLVAVQPKVAFS 2147
QY 2170 RPLQDVVTEKEKVTLECELSRPNVDVRLKDGVELRAGTKMAIAQAQACRSLTIVRCEF 2229
DB 2148 ---EEHPRASSAATARKVKGAPMFLQGLLED--MDLKAGASAFAVGK--LGRKLRPHR--- 2199
QY 2230 ADQGVVYCDADHAQSSASVKVQGRVTYTLIYRVLAEADAGEIOFVAENA-----ESRAQL 2283

```

```

DB 2200 -----STNDADKLAKALAQSLR-----LDEPRASIDSRPESAANAALDEVRAAI 2244
QY 2284 RVKE-----LPVTLVRPLRDKIAMEXHGV--LSCQVVS--RASAQVRWFKSQEQLQPGPKVEL 2337
DB 2245 NGRNKKVCPRKFMVWPKPRKV--LSEYKSLRLKATASGNMPQVHWDEKGIILETCNKYSI 2303
QY 2338 VSDGFLYRKLIIISDVAEDEDITYTCDAGDVKTSAQFFVEEQSITIVRGLQDVVMEPAPAW 2397
DB 2304 YNDGDFYLLVHHVSTFDKGFNCTAAN-----NEGIIIT----- 2337
QY 2398 FICETSIPIPSRPPKMLLKVLOAGNVGLBOGTVHRLMLARTCTSTGPHVFTVHGSR 2457
DB 2338 --CTSEI--DVLENK-----EDSAAQVAKRKSKEAKAFNFIENVLFR 2376
QY 2458 SSARLVVSDIPVVLTRPLEPKTGRELQSVLSCDFRPAPKA--VOMYKDDTLPSP--SEKFK 2515
DB 2377 SQANL-----NESLCVECSVSAYPCASIIWTRNSVLLFOADRYT 2416
QY 2516 MLEBGMALRILRLMPADAGVIRCOA---GSAHSSTEVTVREAREVTVTGPLQDAEATE 2571
DB 2417 MSYDGECSALKFISVTPGDEGTACEAVNELGSAVTNNLOVSG-----VDPNAAE 2467
QY 2572 EGWASFSCELSHEDVEWLSLNGM--PLYNDSPFHEISHKGRHRTLVLKSIQADAGIVRAS 2630
DB 2468 -----GIPPLFR--FEKI-----KSVR----- 2482
QY 2631 SLKVSSTASLEVRVVPVFLKALDLSAERGT-LALQCEVSDPEAHVVRKDGVLQGFSD 2690
DB 2483 --KWDGSRVLEALBV-----QASEP--LQIRLWLNKVTIIVDSP 2518
QY 2691 KYDFLHTAGTGLVVDVSPEDAGLYTCHVCSSETRARVRVHDHLVGTIKLKTMEVLEG 2750
DB 2519 SFSYSRSRNMVFLTIADVPEDGGEYTVKAKNQ-----GIARTMQLDVRNN 2566
QY 2751 E--SCSFEC--VLSHE--SASDPAMVT--VGGTKVSSSRFQATROGRKYILVREAAPSDA 2804
DB 2567 ERSVADEAPRVDFEPTTRSDPGVSVELRAKVIGHDPDMR--NEGDFILRIANVTRADA 2624
QY 2805 GEVFSVRGLTSKASL-----IVRERPAIITKPLEDQWVAGPEDEVLECELSRA 2853
DB 2625 KKYELTAINPQGOANAELELTVQSTKTVGAKPKFNESPIVQ-----TCERNRA 2674
QY 2854 -----GTPV---HML-KDKRAIKSKQYDVVCEGMALVIRGASLKDAGEYTCBEVA 2902
DB 2675 ELRASFSCTPAPACRFVYNGNELIDGLDYITTSDDTNSLLINSVKKHFGELCTIR- 2733
QY 2903 SKSTASLHVEKANCFTTELTNLQVEBKGT-----VPTCKT 2939
DB 2734 -----NQNGEELANAMILSEGECKRPHRIDIVFVCSNS 2765

```

RESULT 15

PN0568

connectin 3B - chicken (fragment)

N:Alternate names: Cn3B protein

C:Species: Gallus gallus (chicken)

C:Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 17-Mar-1999

C:Accession: PN0568

R:Matuyama, K.; Endo, T.; Kume, H.; Kawamura, Y.; Kanazawa, N.; Nakauchi, Y.; Kimura, S.;

Biochem. Biophys. Res. Commun. 194, 1288-1291, 1993

A:Title: A novel domain sequence of connectin localized at the I band of skeletal muscle

A:Reference number: PN0568; MUID:93356802; PMID:8352787

A:Accession: PN0568

A:Molecule type: mRNA

A:Residues: 1-1323 <VAR>

A:Cross-references: DDBJ:DI6541; NID:g391629; PID:d1004495; PID:g391630

A:Experimental source: skeletal muscle

C:Comment: This protein string-like single molecule spans from the Z line to the M line

Query Match

Best Local Similarity 25.3%; Score 869; DB 2; Length 1323;

Matches 298; Conservative 184; Mismatches 528; Indels 168; Gaps 38;

| | | | |
|----|------|--|------|
| Qy | 3969 | SATLVTAAVPVRLRELQHQVEVDGGTAHLCELSEAGASVVEKGSLOLFFCAKYMVQ | 4028 |
| Db | 1041 | TADLTAEAPIQTGSIQNVIVSEHOSATFECEVSFDDAVVTYKGPTELTPSKYSFRS | 1100 |
| Qy | 4029 | DGAABELLVRGVEQEDAGDYTC-----DTGHTQSMASL | 4061 |
| Db | 1101 | EGRCHYMTTHNVTAEDGYSVIARLEPRGEARSTAEI | 1138 |

Search completed: September 13, 2004, 11:39:52
Job time : 411 secs

| | | |
|----|------|---|
| QY | 3012 | EDVVDQEGSSATPRCISIPANYPEVHWF---LDKTPH-----ANELNEIDAQP---3057 |
| Db | 1 | EPFNVLSGENITFTSVKGGSPPLEVKNFWRGSIELAFGHKCNITLQDSVAELBFLDVQPLQ 60 |
| QY | 3058 | -GGYHVLTLRLQALKDSGTYIFEAGDQASALRVTEKPSVFSRELTDATITIGEDLTLV 3116 |
| Db | 61 | SGDY-----TCQVSNEAGKISCTTHLFVKE-PAKFVMKNDLSVEKGNLILE 107 |
| QY | 3117 | CE-TSTCDIPMCWTGDKTLRGSARCOLSHEGHAOLLITGATLQDSGRYKC-----EAG- 3170 |
| Db | 108 | CTVTGTPPSVTVKKNQGVILKHSEKCSITTTSTAILEIPNSKLEDOQOQISCHIENDSQ 167 |
| QY | 3171 | GACSSIVRVHARPVRFQEKDLLEVLGGAAATLRC-VLSSVAAPKYWCYGNNVLRPGDK 3229 |
| Db | 168 | DNCHGAIITL--EPYFVTPLEPVQVTVGDSASLOQVAGTPEMIVSWYKGDTPKLGATAT 225 |
| QY | 3230 | YSLRQEGAMLELVVRNLRPQDSGRYSC---SFGDOTTSATLTV--TALPAQFIGKLRNK 3283 |
| Db | 226 | VRHFKNQVATLVPSQVSDDSGEYICKVENTVGEATSSLLTVQERKLPSPSTRKLDRV 285 |
| QY | 3284 | EATEGATATLURCELSKTAPE--WRKGSETLRDGDRYCURLQDQGAMCELOIRGLAMVD---3338 |
| Db | 286 | HETVGLPVPFDCGIAGSEPEVSWFK-----DNVRVKEDYVNHVTSFTDINVAIDLQILK 337 |
| QY | 3339 | -----AAEYSCVCGEERTSAS-----LTIRPMPAHFIGRLRHOESIEGATATLURCELSK 3387 |
| Db | 338 | TDKSLMGQYCTCTASNAIGTASSGKVLTEGKTPPTFPDTPITPVDGIIGESADFCHISG 397 |
| QY | 3388 | AAP--VEWRKGRBSLRDGRHSLRQDQAVCELOICGLAVADAGEYSCVC---GEERTSA 3441 |
| Db | 398 | TQPIRTWAKDNQBIRTGNGYQISYVENTAHLTILRVDGRSGKYTCYASNEVGKDSCTA 457 |
| QY | 3442 | TLTVK--ALPAKTEGLR-----NEEAVEGATAMLWCELSKVAPEVRWKGPENLRDG 3491 |
| Db | 458 | OLNVKERKTPPTFRKLSEAVEETEGNELKLERVAG-----SPLTVSHYKKNQEVHSS 512 |
| QY | 3492 | DRYILRQEGTRCELOICGLAMADAGEYLCVCGOERTSATILTRAL-----PARFIEDVK 3545 |
| Db | 513 | PHCEISPKNNTLILHIKSVQSDAGLYTCRVNSEAGSVLCTSSVWIREPKKPVFPDQLQ 572 |
| QY | 3546 | NQAREGATAVLOCELSNAP--VEWRKGSETLRDGRYSLRQDGTKCELOIRGLAMVDT 3603 |
| Db | 573 | PAATEGDTLQLSCHVRGSPRIQWLKAGRETRASERCFSFSPANGVALLELAAYTKSDS 632 |
| QY | 3604 | GEYSC-----VCGOERTSAMLTVR---ALP-----IKFTFEGURNBEATEGAT 3642 |
| Db | 633 | GEYVCKASNAGVTDTCRSKVTVEKAALVSAAKKADIEGKLYFVSEPOSIKVVEKTV-AT 691 |
| QY | 3643 | AVLRCELSKNAPVEVWKGH--ETLRDGRHSLRQDQARGCELOIRGLVAEDAGEYLCWC---3698 |
| Db | 692 | FIKVGDDPFPNVKRWKQVQLNQGRIIIOQGDKALEIKDTKTDSGLYKCVAFNQ 751 |
| QY | 3699 | -GKERTSAMLTVRAMPKFI-EGLR-----NEATEGDTATLWCELSKAAPVEWRK- 3747 |
| Db | 752 | HGETSVNLQVBERKQEVVEEDVRGKLRIPTKKKDEDEQTDIDLELLKNVDPKMEYKY 811 |
| QY | 3748 | -----GHETLRDGRHSLRQDGRCELOIRGLAVVDAGEYSCVCG--QERT 3791 |
| Db | 812 | ARMYGITDFRGLLOAFELL----KQREBSHRLEIELTEKAQKEDQFEELVAFIQORL 867 |
| QY | 3792 | SATLTVRALPARFIEDVKNOAREGATAVLOCELSKAAP---VEWRKGSETLRGGDRYSL 3848 |
| Db | 868 | TQT-----EPVTLIRDIENTQVLDEDAIFECETIKINYPIKLSWYKGTOKLQSSDDKYI 922 |
| QY | 3849 | RQDGTGRCELOIHGLSVADTGEYSCVCGOERTSATLTVRAPQVFPREPLQSLQABEGSTAT 3908 |
| Db | 923 | KIEGDRHLIKIKNCCLEDOGNRYIVCGPHIASARLTV--IEPAVERHLHDTTFKEGNTCT 980 |
| QY | 3909 | LQCELSBPTATVWVSKGLQLOANGREPRLOGQTAEVLVLODLOREDTGEVYTCGSGQAT 3968 |
| Db | 981 | LSCOFSTIPNAKSOVWRNGRPKIKIGRYSTQVSDKVHKLIIKNDVTEQOQVYTCCKLDNLET 1040 |

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:05:23 ; Search time 82 Seconds
(without alignments)
5059.695 Million cell updates/sec

Title: US-10-077-130-5

Perfect score: 41273
Sequence: 1 MDQFSGAPFLTRPKAFV.....RNREKRALLYKRNLAQVR 7968

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 2486 | 6.0 | 6632 | 1 UN89_CABEL | O01761 caenorhabdi |
| 2 | 1074 | 2.6 | 4391 | 1 PGBM_HUMAN | P98160 homo sapien |
| 3 | 1023 | 2.5 | 1914 | 1 KMLS_HUMAN | Q15746 homo sapien |
| 4 | 1012 | 2.5 | 1906 | 1 KMLS_CHICK | P11799 gallus gall |
| 5 | 909 | 2.2 | 3038 | 1 TRIO_HUMAN | O75982 homo sapien |
| 6 | 767 | 1.9 | 3707 | 1 PGBM_MOUSE | Q05793 mus musculu |
| 7 | 645.5 | 1.6 | 1176 | 1 KMLS_BOVIN | Q28824 bos taurus |
| 8 | 622 | 1.5 | 1147 | 1 KMLS_RABIT | P29294 oryctolagus |
| 9 | 586 | 1.4 | 1270 | 1 MYPC_MOUSE | O70468 mus musculu |
| 10 | 572.5 | 1.4 | 1709 | 1 SN_HUMAN | Q9b222 homo sapien |
| 11 | 571 | 1.4 | 1274 | 1 MYPC_HUMAN | Q14896 homo sapien |
| 12 | 532 | 1.3 | 1694 | 1 SN_MOUSE | Q82230 mus musculu |
| 13 | 528.5 | 1.3 | 1131 | 1 MYPF_CHICK | P16419 gallus gall |
| 14 | 520 | 1.3 | 1142 | 1 MYPF_HUMAN | Q14324 homo sapien |
| 15 | 519 | 1.3 | 3375 | 1 UN52_CABEL | Q06561 caenorhabdi |
| 16 | 506.5 | 1.2 | 1271 | 1 MYPC_CHICK | Q90688 gallus gall |
| 17 | 499.5 | 1.2 | 607 | 1 KML2_RABIT | P07113 oryctolagus |
| 18 | 495.5 | 1.2 | 595 | 1 KML2_HUMAN | O9h1r3 homo sapien |
| 19 | 489 | 1.2 | 448 | 1 DAK3_MOUSE | O84784 mus musculu |
| 20 | 489 | 1.2 | 448 | 1 DAK3_RAT | O88764 rattus norv |
| 21 | 480.5 | 1.2 | 609 | 1 KML2_RAT | P20689 rattus norv |
| 22 | 475.5 | 1.2 | 370 | 1 DAK2_MOUSE | Q8vdf3 mus musculu |
| 23 | 471 | 1.1 | 1432 | 1 DAK1_HUMAN | P53355 homo sapien |
| 24 | 469 | 1.1 | 1442 | 1 DAK1_MOUSE | O80ye7 mus musculu |
| 25 | 467 | 1.1 | 1141 | 1 MYPS_HUMAN | Q08072 homo sapien |
| 26 | 466.5 | 1.1 | 454 | 1 DAK3_HUMAN | Q43293 homo sapien |
| 27 | 465.5 | 1.1 | 370 | 1 DAK2_HUMAN | Q9uik4 homo sapien |
| 28 | 464.5 | 1.1 | 1050 | 1 ULK1_HUMAN | O75385 homo sapien |
| 29 | 462.5 | 1.1 | 2012 | 1 DSCA_HUMAN | O60469 homo sapien |
| 30 | 460.5 | 1.1 | 438 | 1 KMLS_SHEEP | O02827 ovis aries |
| 31 | 456.5 | 1.1 | 1465 | 1 MYM2_HUMAN | P54296 homo sapien |
| 32 | 451.5 | 1.1 | 1051 | 1 ULK1_MOUSE | O70405 mus musculu |
| 33 | 451.5 | 1.1 | 1451 | 1 MYM1_HUMAN | P52179 homo sapien |

ALIGNMENTS

RESULT 1

```

UN89_CABEL
ID UN89_CABEL STANDARD; PRT; 6632 AA.
AC O01761; Q17362;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT assembly, encodes a giant modular protein composed of Ig and signal
RT transduction domains";
RL J. Cell Biol. 132:835-848(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U33058; AAB00542.1; -.
CC DB; AF003131; AAB54132.2; -.
CC PDB; 1FHO; 20-DEC-00.
CC WormPep; C09D1.1; CE30426.

```

```

34 451 1.1 371 1 S17B_RAT Q91xs8 rattus norv
35 447.5 1.1 397 1 S17A_RABIT Q9gm70 oryctolagus
36 445.5 1.1 414 1 S17A_HUMAN Q9uee5 homo sapien
37 441.5 1.1 374 1 KCCL_EAT Q63450 rattus norv
38 440 1.1 424 1 KPSH_HUMAN P11801 homo sapien
39 439.5 1.1 1450 1 MFSF_CHICK Q02173 gallus gall
40 438.5 1.1 372 1 S17B_MOUSE Q8B948 mus musculu
41 437.5 1.1 374 1 KCCL_MOUSE Q81ys8 mus musculu
42 432 1.0 794 1 K111_HUMAN Q8tdc3 homo sapien
43 428 1.0 372 1 S17B_HUMAN O94768 homo sapien
44 427 1.0 370 1 KCCL_HUMAN Q14012 homo sapien
45 426 1.0 1666 1 MYM1_MOUSE Q62234 mus musculu

```

DR InterPro; IPR008957; FN III-like.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003598; IG c2.
 DR InterPro; IPR003006; IG_MHC.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR007850; RCSD.
 DR InterPro; IPR000219; RhoGEF.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00041; fn3; 1.
 DR Pfam; PF00047; ig; 47.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF01177; RCSD; 5.
 DR Pfam; PF00621; RhoGEF; 1.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00408; IGC2; 23.
 DR SMART; SM00325; RhoGEF; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS0010; DH 2; 1.
 DR PROSITE; PS0835; IG-LIKE; 49.
 DR PROSITE; PS00003; PH_DOMAIN; 1.
 DR PROSITE; PS00002; SH3; 1.
 Kw Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
 3D-structure.
 KW DOMAIN 63 127 SH3.
 FT DOMAIN 152 330 DH.
 FT DOMAIN 342 498 PH.
 FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
 FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
 FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
 FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
 FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
 FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
 FT DOMAIN 1272 1315 THR-RICH.
 FT DOMAIN 1375 1475 RCSD 1.
 FT DOMAIN 1479 1585 RCSD 2.
 FT DOMAIN 1597 1695 RCSD 3.
 FT DOMAIN 1700 1799 RCSD 4.
 FT DOMAIN 1800 1860 RCSD 5.
 FT DOMAIN 1882 2067 IG-LIKE C2-TYPE 7.
 FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
 FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
 FT DOMAIN 2367 2359 IG-LIKE C2-TYPE 10.
 FT DOMAIN 2455 2455 IG-LIKE C2-TYPE 11.
 FT DOMAIN 2564 2564 IG-LIKE C2-TYPE 12.
 FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
 FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
 FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
 FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
 FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
 FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
 FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
 FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
 FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
 FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
 FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
 FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
 FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
 FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
 FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
 FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
 FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
 FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
 FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
 FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
 FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
 FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
 FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
 FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
 FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
 FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
 FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
 FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.

5487 5578 5595 5685 5701 5790 5815 5904 5925 6014 6038 6130 6239 6368 6413 6502 6596 6507 6568 6578 6582 6589 6596 6604 6614 6624 6634 6644 6654 6664 6674 6684 6694 6704 6714 6724 6734 6744 6754 6764 6774 6784 6794 6804 6814 6824 6834 6844 6854 6864 6874 6884 6894 6904 6914 6924 6934 6944 6954 6964 6974 6984 6994 7004 7014 7024 7034 7044 7054 7064 7074 7084 7094 7104 7114 7124 7134 7144 7154 7164 7174 7184 7194 7204 7214 7224 7234 7244 7254 7264 7274 7284 7294 7304 7314 7324 7334 7344 7354 7364 7374 7384 7394 7404 7414 7424 7434 7444 7454 7464 7474 7484 7494 7504 7514 7524 7534 7544 7554 7564 7574 7584 7594 7604 7614 7624 7634 7644 7654 7664 7674 7684 7694 7704 7714 7724 7734 7744 7754 7764 7774 7784 7794 7804 7814 7824 7834 7844 7854 7864 7874 7884 7894 7904 7914 7924 7934 7944 7954 7964 7974 7984 7994 8004 8014 8024 8034 8044 8054 8064 8074 8084 8094 8104 8114 8124 8134 8144 8154 8164 8174 8184 8194 8204 8214 8224 8234 8244 8254 8264 8274 8284 8294 8304 8314 8324 8334 8344 8354 8364 8374 8384 8394 8404 8414 8424 8434 8444 8454 8464 8474 8484 8494 8504 8514 8524 8534 8544 8554 8564 8574 8584 8594 8604 8614 8624 8634 8644 8654 8664 8674 8684 8694 8704 8714 8724 8734 8744 8754 8764 8774 8784 8794 8804 8814 8824 8834 8844 8854 8864 8874 8884 8894 8904 8914 8924 8934 8944 8954 8964 8974 8984 8994 9004 9014 9024 9034 9044 9054 9064 9074 9084 9094 9104 9114 9124 9134 9144 9154 9164 9174 9184 9194 9204 9214 9224 9234 9244 9254 9264 9274 9284 9294 9304 9314 9324 9334 9344 9354 9364 9374 9384 9394 9404 9414 9424 9434 9444 9454 9464 9474 9484 9494 9504 9514 9524 9534 9544 9554 9564 9574 9584 9594 9604 9614 9624 9634 9644 9654 9664 9674 9684 9694 9704 9714 9724 9734 9744 9754 9764 9774 9784 9794 9804 9814 9824 9834 9844 9854 9864 9874 9884 9894 9904 9914 9924 9934 9944 9954 9964 9974 9984 9994

QY 420 -GNLLRLPRKTAVRAGDTAMFCVELAVPGP---VHWRNQBEEVAGG-RVAISAGTR 474
Db 945 APTFLKJEDQT-VKTGEFAVF---ETTVRGPNPEVTFWINGHMDQSGSPVKLEAHND 1001
QY 475 HTLTISQCCLEEDVGOVAFMA---GDCOTSTRFCVSPAPRPPLOPPVPPVKARMESVVI 530
Db 1002 HKLTIDSA--QVAGTVLCRAENAVGRFETKARLVLAPEKQKPPKFVEILVDKTE--- 1055
QY 531 LSWSPPPHGERPVTLDGVLVEKKLGTVTWIRCHEAEHWATPELTADVABEGNQFRVS 590
Db 1056 -----TVNDTVW-----FEVRVE 1068
QY 591 ALNSFGQSPYLEFFGCTVHLAPKLAVRTPLKVAQAVEGGEVTFVSDLTVASAGEWFLDQQA 650
Db 1069 G-----EPKPTV-----TWYXGEE 1083
QY 651 LKASSVYIHCDRPHRTITIREVPASLHGAQLKFWANGIESIRMEVRAAPCLTANKPPA 710
Db 1084 LKQSDRVIREBFDGSIKISIKNIKIEDAG-EIRAVATNSEGSDBTAK---LTVQKPF 1138
QY 711 AA---AREVTLARLHEEAQLLAELSQQAAAATWTKDGRTLSPGPKEVQASAGRVLVLRD 767
Db 1139 APEPDLRFVSLTEKSE--AVFSAHAFGI-----PLTYEWSVN-GRKV----- 1180
QY 768 VARDAGLYECVSRGRRIAYQLSVQGLARFLHKDAGSCVDVAGGPAQFECETSEAHVH 827
Db 1181 ---RDG-----QEGARVTRDESTVDGASILTIDTA----- 1207
QY 828 VHWYKDMGELGHSGERFLQEDVGTHRLVAATVTRQDEGTYSCEVGSVDVFLRVSEPK 887
Db 1208 -TYKS---EVNH-----ITISVVAENT-----LGAETGAQUTI-EPK 1240
QY 888 ---VYFAKEQLARRKLOAEAG-----ASATLSCEVAQAQTEVTWYKDGKGLSSSK 935
Db 1241 KESVVVEKQDLSSEVQXEAQVKEASPEATITIMETSLSKTITMSTTEVTSVGG 1300
QY 936 VCMEATGTRRLVVOQAQADAG--EYSCAAG---QRLSFHLDVKE--PK--VYFAKD 985
Db 1301 VTVETKESESATTVIGGSGGVTEGSIYSVKIEVSKTDSQTDVREGTPKRVSFAEE 1360
QY 986 QVAHSEVAQAEAGANATLSCEVAQAQAEVWYKDG-----KGLS 1023
Db 1361 ELPKEVIDSDRKKKSPSPDKKESPEKTEKPAKSPKTKTGEEVKSPEKSPASPTKEX 1420
QY 1024 SSLKVHVEAKCRRRLVVOQAGKTADGYSCEAGQORVSFLHITPEKMPFAKQOS--- 1079
Db 1421 SPAAEEVKSPTKKEKSPSPSTKKEKSPSPSTKKTGDYKKEK---SPPKSPTKKEKSPK 1477
QY 1080 -----VHNEVQAEAGASAMLSCEVAQAQAEVWYKDG-----GKLS 1122
Db 1478 EDVKSPEKSPDNTNIVESSSETTIEKTTTENTHESSESRTSVKKEKTEPKVDE 1537
QY 1123 EVKGTCLRLVLPQAGKADAGEYSCEAGQORVSFHLHITP-KGVFAKEQSVHNEVQABAG 1181
Db 1538 KPXSPTKDKGSPKS-----ITEEIKSPVKKEKSPKEVEKPAS 1576
QY 1182 TTAMLSCEVAQAPOTEVWYKDGKGLSSSKVYMEVKGCTR-----RLVVOQV----- 1228
Db 1577 PTK-----KEKSP-----KPASPTKGENEVKSPTKKEKSPKSVVEELKSPKES 1623
QY 1229 -GKADAGEYSCEAGQORVSFQHLHITPEKPAVFAKQVHNEVTRTEAGASATLSCEVAQAQ 1287
Db 1624 PEKAD-----DKPKSPKKEKSPKESA-TEDVKSPTKK-EKSPEK 1662
QY 1288 EVTWYKDGKGLSSSKVRIEBAAGCMRQVUVQAGQADAGEYTCAGGQORLSFHLVDSEPK 1347
Db 1663 EEKPTSPTKGSSPTKKTDE-----VKSPTKKEKSPQTV-----EKPA 1702
QY 1348 AVFAKEQLAHRKVOAEAGAIATLSCEVAQAQTEVTWYKDGKGLSSSKVYMEVKGCTRRL 1407
Db 1703 SPTKKEKSPKSVVEEVKSPEKSPKAEKPK-----SPTKKEKSPKSAAEVKSPTK- 1757

QY 1408 VVOQAQADTGEYSCEAGGQORLSFSDVABPKVVF-AKQOPVHREYQAOAGASTTSLCEV 1466
Db 1758 --KEKSPKSAEBKPSPTKKGSSPVKMADDEVKSPKKEKPEKVEEPASPTKKEKTP 1815
QY 1467 AQOQTEVMYKQKGLKLSFSKVRMEAVGTRRLVVOQAQADAGEYSCEAGQORLSFHLH 1526
Db 1816 EKSAABELASPTKKEKSPSPSTK-----KTGDESKEKSPKPS----- 1853
QY 1527 VABPKAVFAKEQP-----ASREVAEAGTATSLCE-----VAQAOT 1563
Db 1854 -EKPSPPTKSPGSPGPKKKSKSPEAEXPAPKLTRODKLQTVNKTDLAHEVTVVEHAT 1912
QY 1564 EVTWYKDGKGLSSSKVYMEVAGVCTRRLLVVOQAQADAGEYSCK----- 1607
Db 1913 ECKWFLDGXEITTAQGVTV-----SKDDQEFRCSDITTMFGSGTVSVWAS 1958
QY 1608 --AGDORLSFHLHVAEPKVVFAKEQPAHREVAEAGASATLSCEVAQAQAEV-TWYKDGK 1664
Db 1959 NAAGSVETTELKVLSTPKETKKEPFTDLRDMVETKQDVTQVMDVJALHSPLYKWTQNGN 2018
QY 1665 KLSSSSK-VRVBAVGTRRLVVOQAQADAGEYSCEAGQORLSFRLHVALEPQISERPC 1723
Db 2019 LLEDGKNGVTIKNEENKSLIIPNA--QDSGKITVEASNE---VGSSESAQLTVNPP 2071
QY 1724 RREPLVV-----KEHEDIILTATLPSTAATVTLWLDGVBIRSKKHETASQODTHT 1775
Db 2072 STTPIVVDGPKSVTIXETETAETAFKATISGFPAPTAKWT-NEKIVEESRTTIITKTEDVVT 2131
QY 1776 LTVHGAQVLDLSAIYSCRVAEGQDFPVQVEEVAARCLLEPVCGLGCTVTLACLSPA 1835
Db 2132 LKISNAKI-----EQTGTVKVTAQNS-- 2152
QY 1836 CAEVMWRCNGTQPRVKRFQMVAGPVRSLTVGLRAEDAGEYVCESRDHDTSAQLTVSV 1895
Db 2153 -----AGQ---DSKQADLKVEPNVKA 2170
QY 1896 PRVVKEMSGLSTVWABEGGEATFOCVV-SPS-DVAVVWFRDGLALOPSEKFAISQSG-AS 1952
Db 2171 P---KPKSOLTDKVADEGEPLRWNLDELGPSPGTEVSWLLNQPLTKSDTVQVVDHGDGT 2227
QY 1953 HSLTISDLVLEDAQOITVEABEGAS-----SSAALRVEAPVLFKKLE-----PQ-----TVE 2000
Db 2228 YHVTVIAEAKPEMSGTLTAKAKNAAGBCEKSAKVTYNGG---NKKPEFVQAPQNHETTL 2283
QY 2001 ERSSVTLVELT-RPMPBLRWTRNATALAPGNVE---THAEG-----ARHLVLHNVGF 2051
Db 2284 E--SVKFSATVTKGMPNVTVYLNKKLQSEBEVKVYVHETGKTSIRIQKPLMEHN--- 2338
QY 2052 ADRGFFGCCTP-----DDXTQA-KLTIVEM--QVRLVRGLQAVEARE-QGT 2093
Db 2339 ---GTRIRVEAENVSGKVQATAQLKYDKTEVPKFTTNDROVK-----EGEDVKFT 2387
QY 2094 ATNEVOLSHADVGSWTRDGLRFQOQPTCHLAVRGPMTHTLTLGURPEPSGLMVFKAEG- 2152
Db 2388 ANVE---GYPEPSVAMTLNGEPVSKHPNTVTVDKQGEHTIEISAVTPEQAGLSCEATNP 2444
QY 2153 VHTSARLV-----VTELVPFSRPLQDVVTTKEKVTLECSLS--RPNVDVRLWKDGYE 2204
Db 2445 VSGKEDVQLAVKVGADAP--TFAKNLEDLITIEGELTMDAKNLVKKPKTWTWKDGYE 2503
QY 2205 LRA-GKTMIAAQAQACRSITTYRCFPADQGVVVCDAHD-----AQSSASVYKQRTYTL 2259
Db 2504 ITSDGHYKIVEBEDGSLKSLILQTKLEDKGRITKAESBFGVAECASLGV-----VK 2556
QY 2260 RRVLAEDAGEIQFVAENAESRAQLRVKELPVTVLPRDLRKIAMKHGVLQCV---SRA 2316
Db 2557 GRPMKAPA---FQSDIA-----PINLT-----EGDTLECKLITGDP 2590
QY 2317 SAQVRWFKGSQELQPGPKYELV-SDGLYKLLISDVHAEDEDTYTC---DAGDVKTSQA 2371
Db 2591 TPVFWYIGTQLVCATEDEISNANGVY-TMKIHGVTADMTGKICVAYNKAGEVSTEG- 2648
QY 2372 FVFEOSIITVRLGQDVTVNEPAPAFWE---CETSI-----PSVRPPK 2411

QY 4268 SFQCRLSRASGOEARWALGVPLQANEMNDITVEQTLHLTLHLKVLTELDAGTVSFHV-- 4325
D 4798 VFEVKI--QGEPEDEVRLDANVISAGANAIIKIDDTTYRLIIISADLKXAGEVTVVEIN 4856
QY 4326 --GTCSSBAQKVTAKNTVRLGLNVALEGEALFECQLSQPEVAHAHTWLLDDPEVRTS 4383
D 4857 ESGKASDAKGEVDEKPEIPEGLNIEIPEGGDDVFKVEVSAP--VRQVKYKNDQEIKN 4915
QY 4384 ENAEV-----VFENGRLHLLKMLRQDSCRVTFLAGDVVTSAFLT--VR 4428
D 4916 SHLEAKKIPKKYELAINRAQLDGDADYKVVLSN-----AAGDCSSAALTIVK 4964
QY 4429 GWRLEIPLKVAIVRAGAARFCTTLSEAVPVEGASWYINGAAVQDDSDMTVTADGSH 4488
D 4965 PNVLKIDGLKDVDVEEPQVELKVKV--EGIP--KVIKWKNGQELKPD-----ADG-- 5013
QY 4489 QALLRSQAHPHAGEVTFACDASARLTVLGLPDPEDAEVVAHSSHTVTLSSWAAPMS 4548
D 5014 -----FKPEEPESGE-----FSLTIP--SSKKS 5035
QY 4549 DGGGLCGYRVEVEKAGATGOWRLCHELVPGCEVVDGLAPGETYRFRVAAGVPVAGEPV 4608
D 5036 DGG--AYR-----VVLGNDKGEVY-----SGSVV 5057
QY 4609 HLPQTVRLAEBPKVPPOP--SAPESQVAAAGEDVSLELVVAA--GEVIVHKGMERIQP 4665
D 5058 H-----VKSASEPTSGANFLSPDKTEVEEGDMLTLQCTIAGEPFPPEVWEXDGVVLQK 5113
QY 4666 GGRREV--VSQGRQOMLVTKGTABDQGEYHCGLAQGSICPAAATFOVALSPASVDEAPOP 4724
D 5114 DDTITMRVALDGTATLRSKAKSDIGQYRV--TAKNEAGSATSCKTVT-----BQGEOP 5168
QY 4725 SLP-----PEAAQEGDLHLLEALAKRMSREPTL-----DSISELPEED 4765
D 5169 SKPKFVIPLEKTAALPGDK-----KEFNKVRGLPKPTLQWFLNGIPIKFDRIITLDMAD 5224
QY 4766 GRSQRLPOEAEEVAPDLSEGYSTADELARTGDADLSHTSDDESACTPSLTVILKAGR 4825
D 5225 GNYCLTIRDVRE-----EDFTLTKTAK-----NENGIDE-----TVCEPQQAGH 5265
QY 4826 PGTSLASKVGAPAPSVKPOQOQEPALAAVRPPLGLDLSKDLGDPMSDKAAVKITQAAAFK 4885
D 5266 -----DGSRDDLRYP----- 5276
QY 4886 YKVRKEMKQOQPMFSTFGTEAQVGDALBLECVKASKADVRARWLKDGVELTDGRHHH 4945
D 5277 -----PRFNVLWDRRIPVGDPMFIECHVDANPTAEVWFKDGKKEHTAHE 5324
QY 4946 IDQLGDGTCSLIAGLDRADAGCYTCQVSNKFGQWTHSACVWVSGSESEAESGGELDD 5005
D 5325 IRNTVDGACRIKIIPFEESDIGVMCVAVNELGQAEQCATVQVILEHVEEKR----- 5378
QY 5006 AFRAARLRLHRTKSPAESVDSBELFLSADGPAEPPEADWQTYREDEHFCIRFEAL 5065
D 5379 --REYAPKNPLEDKTVNGQPIRLSKVD--AI PRASVW--YKDG-----LPLR 5424
QY 5066 TEACAVTRFOEMFATLIGVIEIKLVEGPRRVMC--SK-----ETPAPVVPPEPLPSL 5119
D 5425 ADSRTSIQYEDGCTATLAINSTE--EDIGAYR--CVATNAHGINTSCSVNVKVPQEV 5480
QY 5120 LTSDAAPVFLTELQNEQVQDGPVSVDCVVTGQPMPSVRWPKDGLLEEDHVMINEDQ 5179
D 5481 KKEGEEPFRTKGLVDLWADRGSFTLKCAVTGDPFPEIKWYRNGQLLRNGPRTVIETSPD 5540
QY 5180 GGHQLIITAVPADGVVECLAENSMGVSTKABLRVDLTSTDYDTAADATESSESYESAQ 5239
D 5541 GSCSTVNNESTWDEGIYRCAENAHGAKAKTOAHVQM-----ALGKTE----- 5585
QY 5240 GYLSREGEGETESTDEQQLPQVVEELRDLOVAPGTRLAKPQLKVKGVGAPRLYFMDQ 5299
D 5586 -----KPKWDECKPKPKFILELSDMSVSLG--NVIDLECKVTGLPNPSVKWSDGG 5633

QY 5300 PLTASAHIRMTGK--KILHTLEIISVTREDSDGOYAAVISNAMGAAYSSARLLVLRGDEPE 5357
D 5634 PLIEDSRFEMSNEASGVVQLAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAK 5687
QY 5358 EXPASDVHQLVPPRMLERFTPKVKKSSITFSVKVGRPVPVTHWLEBEA-----ERG 5412
D 5688 DGLSGVVTASOPFRFTLXMGDVRTTEGQPLKLECKVDASPLPEVMWYXGALVTPSDRI 5747
QY 5413 VLMIGDPTPGYTVASSAQOHSILVLDVGRHQHGTYTCTIASNAAGALCSASLHVSLPK- 5471
D 5748 QISLSPD---GVA-----TLLIPSCVYDDGIRVVIATNPSGTAQDGTATVKLPRD 5797
QY 5472 -----VEQEYKVEALIS-----TFL 5487
D 5798 SGARRSADRDVEDANKAPKLEPLENIRIPEXQSFRLCRCKFGCDPKPTIKWFKDGERVFP 5857
QY 5488 QGTTOAI--SAQG-----LETASFADLGQORKEEPLAAKEALGHLSLAEVTEFLQKLS 5541
D 5858 YGRLOLIESPDGCELVVDSATQDAGYR--CVAENTYG----- 5895
QY 5542 QITEMVSAKITAKLQVPGD--SDSDSKTSPASPRHGRSRPSSSIQSSSESDGDAR 5598
D 5896 -----SAR--TSCDVNVIRDRKPRDIDS-----SIREGKA--PGFTTPTLIRRAKPGDS- 5941
QY 5599 GEIPDIYVVTADYLPGLAEQDAITLREGQYVEVLDAAHPLRLVTRTKTKSSPSRQGWYS 5658
D 5942 -----VTFECLPFGNPPFSI-----KWL-----KDG--L 5963
QY 5659 PAYLDRRLKLSPEWGAEEAPEFPGEAVSDEYKARLSSVIOELLSEQAQFVBELOLQSH 5718
D 5964 ELFSDSEKIM-----EAAADG-----TQRLILSDVTFLESEGYF-- 5996
QY 5719 HLQHLERC-----PHVPIAVAGOKA--VIFRNVDDIRGFHSFLOELQOCCDDDDVAMCFK 5773
D 5997 -----RCVATNEH-----GTASTKAEVLIEGDRITIG----- 6022
QY 5774 NQAAPEQYLEFLVGRVQAESVVVSTAIQEFYKYAEALLAGDPSOPPPPLQHYLEQPV 5833
D 6023 -----SRPLP----- 6027
QY 5834 ERVQRYQALLKELIRKARNQNCALLQOAYAVVSALPORAENKLHVSIMENYPTGLEAL 5893
D 6028 -----EYN 6030
QY 5894 GBPIROGHFIVMEGAPGARPKHGNHRVFLFRNLHVICKPRDSDTDTVSVFRNMKL 5953
D 6031 GBPEE-----CKPR-----IRRGLYNM 6047
QY 5954 SSIDLNDQVEGDDRAPEVWQEREDSVRYKYLQARTAIKSSWVKEICIGIQORLALPVMRP 6013
D 6048 S-----IHE----- 6051
QY 6014 PDPEELADCTAELGETVKLACRVCTGTPKPVISWYKDGKAV--QVDPHILIEDPDGSC 6070
D 6052 -----GNVEMIVCATGIPPTTKWYKDGQEI VGDGPKRVIPTDGRH 6097
QY 6071 ALILDSLGTGDSQVMCFASAGNCSTLGLKILV-----QVPPRFVNVKVR 6115
D 6098 HLIVNASPDDEGEYSLEATNKLGSAKTEGSLNIIRPRIADADERGMPFPFPGVRQLK 6157
QY 6116 ASPFVEGDAQCTCTIEGAPYQIRWYKDGALLTTGNKFPQTLSEPRSGLLVLVIRAAK 6175
D 6158 NKHFVFNHPTI FDLVWGHGPAPEVFNHNGKIVFGRIK--IQSCGGGSHALIILDTLE 6216
QY 6176 DLGLYECELVNLGSGARASAEIRIQSPMLQAEQCHREQLVAAVEDTTLERADQEVTSVL 6235
D 6217 DAGEYATAKNHSAGASSAVLDVTVPFL-----DSIKFNGEIDVTPLY 6260
QY 6236 KRLLG-----PKAPGSTGLTGP-----PCPRGAP----- 6262
D 6261 TEEYGFKLNTASLPTFPDRGPFKEVTHYLTLSWIP--KRAPPYQVTVYVIEIRELPE 6320
QY 6263 ---ALQETOSQPPVT-----GISE--APAVPP-----RVPQPLLHE 6293

RT "Myosin light chain kinase in endothelium: molecular cloning and
 RL regulation.";
 RA Am. J. Respir. Cell Mol. Biol. 16:489-494 (1997).
 RN [2]
 RP REVISIONS.
 RA Birukov K.G., Garcia J.G.N.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3A; 3B AND 4).
 RC TISSUE=Umbilical vein;
 RX MEDLINE=99216419; PubMed=10198165;
 RA Lazar V.L., Garcia J.G.N.;
 RL "A single human myosin light chain kinase gene (MLCK; MYLK).";
 RN Genomics 57:256-267 (1999).
 RL [4]
 RP REVISIONS (ISOFORM 2).
 RA Birukov K.G., Garcia J.G.N.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 923-1914 FROM N.A.
 RC TISSUE=Hippocampus;
 RX MEDLINE=96121365; PubMed=8575746;
 RA Potier M.-C., Chelot E., Pekarsky Y., Gardiner K., Rossier J.,
 RL Turnell W.G.;
 RT "The human myosin light chain kinase (MLCK) from hippocampus:
 RT cloning, sequencing, expression, and localization to 3qcen-q21.";
 RL Genomics 29:562-570 (1995).
 RN [6]
 RP SEQUENCE OF 1614-1914 FROM N.A.
 RC TISSUE=Lung, and Placenta;
 RX MEDLINE=20007838; PubMed=10536370;
 RA Waterson D.M., Schavocky J.P., Guo L., Weiss C., Chlenski A.,
 RA Shinsky V.P., Van Eldik L.J., Haiech J.;
 RT "Analysis of the kinase-related protein gene found at human chromosome
 RT 3q1 in a multi-gene cluster: organization, expression, alternative
 RT splicing and polymorphic marker.";
 RL J. Cell. Biochem. 75:481-491 (1999).
 RN [7]
 RP SEQUENCE OF 1456-1914 FROM N.A.
 RC TISSUE=Placenta;
 RA Waterson M.D.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
 CC SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
 CC IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
 CC FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS. CENTRAL
 CC DETERMINANT IN THE DEVELOPMENT OF VASCULAR PERMEABILITY AND TISSUE
 CC EDema FORMATION. IN THE NERVOUS SYSTEM IT HAS BEEN SHOWN TO
 CC CONTROL THE GROWTH INITIATION OF ASTROCYTIC PROCESSES IN CULTURE
 CC AND TO PARTICIPATE IN TRANSMITTER RELEASE AT SYNAPSES FORMED
 CC BETWEEN CULTURED SYMPATHETIC GANGLION CELLS. CRITICAL PARTICIPANT
 CC IN SIGNALING SEQUENCES THAT RESULT IN FIBROBLAST APOPTOSIS.
 CC -!- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
 CC light-chain] phosphate.
 CC -!- SUBUNIT: ISOFORM TELOKIN BINDS CALMODULIN.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=6;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1; Synonyms=Non-muscle isozyme;
 CC IsoId=Q15746-1; Sequence=Displayed;
 CC Note=The smooth muscle isozyme and telokin are produced by
 CC alternative initiation at Met-923 and Met-1761 of isoform 1;
 CC Name=2;
 CC IsoId=Q15746-2; Sequence=VSP_004791;
 CC Name=3A;
 CC IsoId=Q15746-3; Sequence=VSP_004792, VSP_004794;
 CC Name=3B;
 CC IsoId=Q15746-4; Sequence=VSP_004791, VSP_004792, VSP_004794;
 CC Name=4;
 CC IsoId=Q15746-5; Sequence=VSP_004792, VSP_004793;
 CC Name=Del-1790;
 CC IsoId=Q15746-6; Sequence=VSP_004795;
 CC Event=Alternative initiation;
 CC

CC Comment=3 isoforms, 1/Non-muscle isozyme (shown here),
 CC smooth-muscle isozyme and telokin, are produced by alternative
 CC initiation at Met-1, Met-923 and Met-1761. Telokin has no
 CC catalytic activity;
 CC TISSUE SPECIFICITY: SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES ARE
 CC EXPRESSED IN A WIDE VARIETY OF ADULT AND FETAL TISSUES AND IN
 CC CULTURED ENDOTHELIUM WITH QUALITATIVE EXPRESSION APPEARING TO BE
 CC NEITHER TISSUE- NOR DEVELOPMENT-SPECIFIC. NON-MUSCLE ISOFORM 2 IS
 CC THE DOMINANT SPLICE VARIANT EXPRESSED IN VARIOUS TISSUES. TELOKIN
 CC HAS BEEN FOUND IN A WIDE VARIETY OF ADULT AND FETAL TISSUES.
 CC -!- PTM: MLCK IS PROBABLY DOWN-REGULATED BY PHOSPHORYLATION.
 CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
 CC -!- SIMILARITY: Contains 9 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U48959; AAC18423.2; -
 CC EMBL; AF069601; AAD15921.2; -
 CC EMBL; AF069602; AAD15922.1; -
 CC EMBL; AF069603; AAD15923.1; -
 CC EMBL; AF069604; AAD15924.1; -
 CC EMBL; X85337; CAA59685.1; -
 CC EMBL; X808771; AAD51380.1; -
 CC EMBL; X808771; AAD51380.1; JOINED.
 CC EMBL; AF096766; AAD51380.1; JOINED.
 CC EMBL; AF096767; AAD51380.1; JOINED.
 CC EMBL; AF096768; AAD51380.1; JOINED.
 CC EMBL; AF096769; AAD51380.1; JOINED.
 CC EMBL; AF096770; AAD51380.1; JOINED.
 CC EMBL; AF096771; AAD51381.1; -
 CC EMBL; AF096771; AAD51381.1; JOINED.
 CC EMBL; AF096769; AAD51381.1; JOINED.
 CC EMBL; X808770; AAD51381.1; JOINED.
 CC EMBL; X80870; CAA62378.1; -
 CC HSSP; Q63450; 1A06.
 CC Genew; HGNC:7590; MYLK.
 CC MIM; 600922; -
 CC GO; GO:0004687; F-Myosin-light-chain kinase activity; TAS.
 CC GO; GO:0004688; P-protein amino acid phosphorylation; TAS.
 CC InterPro; IPR008957; FN_III-like.
 CC InterPro; IPR003961; FN_III.
 CC InterPro; IPR007110; IG-like.
 CC InterPro; IPR000719; Prot kinase.
 CC InterPro; IPR008271; Ser_thr_kin_AS.
 CC Pfam; PF00041; fn3; 1.
 CC Pfam; PF00047; ig; 8.
 CC Pfam; PF00069; pkinase; 1.
 CC ProDom; PD000001; Prot_kinase; 1.
 CC PROSITE; PS00835; IG LIKE; 9.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00108; PROTEIN_KINASE_SF; 1.
 CC Transferrase; Serine/threonine-protein kinase; Calmodulin-binding;
 CC ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
 CC Alternative initiation; Alternative splicing.
 CC CHAIN 1 1914
 CC MUSCLE ISOZYME.
 CC MYOSIN LIGHT CHAIN KINASE, ISOFORM NON-
 CC MUSCLE ISOZYME.
 CC MYOSIN LIGHT CHAIN KINASE, ISOFORM
 CC SMOOTH-MUSCLE ISOZYME.
 CC MYOSIN LIGHT CHAIN KINASE, ISOFORM
 CC TELOKIN.
 CC FOR ISOFORM SMOOTH-MUSCLE ISOZYME.
 CC INIT_MET 923 923
 CC INIT_MET 1761 1761
 CC DOMAIN 33 122
 CC DOMAIN 161 249
 CC DOMAIN 414 503
 CC DOMAIN 514 599
 CC DOMAIN 514 599

| | | | | | | | | | | | | |
|-----------------------|----------|--|--|--|-------|--|--|--|-----------------------------------|-------|----------|-------|
| FT | DOMAIN | 620 | 711 | IG-LIKE C2-TYPE 5. | Db | 331 | TAPQTFVLOKTSSTITLOARVQPEPPAPGLGV | -----LSPSGEER | -----KRPAAPPRA | 380 | | |
| FT | DOMAIN | 721 | 821 | IG-LIKE C2-TYPE 6. | Qy | 5113 | --PEPLPSLLTSD | -----AAPVFLTELQNEVQDGYPSFDCVVTGQ | ----- | 5152 | | |
| FT | DOMAIN | 1098 | 1186 | IG-LIKE C2-TYPE 7. | Db | 381 | TFPTPQGLSGDVSVAANRRIPMEGQRDSAPFKFESKPSQOEKVENQTKFRCEVSGI | ----- | ----- | 440 | | |
| FT | DOMAIN | 1238 | 1326 | IG-LIKE C2-TYPE 8. | Qy | 5153 | PMSVRFKDKGLLEEDDHYNEDQOQGHQIITAVVPADMGVYRCLAENSMGVSTKA | ----- | ----- | 5212 | | |
| FT | DOMAIN | 1343 | 1413 | FIBRONECTIN TYPE-III. | Db | 441 | PXEVANFLEGTPVRRQEGSIEVYDAGSHYLCLLKA | ----- | ----- | 500 | | |
| FT | DOMAIN | 1464 | 1719 | PROTEIN KINASE. | Qy | 5213 | ELRVOLTSYDYDAADATESYFSAQGYLSREOEGTSTTDEQQLQVQVVEERDLQVA | ----- | ----- | 5272 | | |
| FT | DOMAIN | 1711 | 1774 | CALMODULIN-BINDING. | Db | 501 | TIQVERLA | -----VMEVAPSFSS | ----- | 526 | | |
| FT | DOMAIN | 1809 | 1898 | IG-LIKE C2-TYPE 9. | Qy | 5273 | PGTRIAKQLKVGYPAPRLYWFQDQGLTASAHIRMTGKXILHTLEIISVTRDSGOYA | ----- | ----- | 5332 | | |
| FT | NP_BIND | 1470 | 1478 | ATP (BY SIMILARITY). | Db | 527 | EQCDFV-LQCSVRGTFVPRITWLLNGQPI | -----QYARSTCEAGVAELHIQDALPEDHGTYT | ----- | 582 | | |
| FT | BINDING | 1493 | 1493 | ATP (BY SIMILARITY). | Qy | 5333 | AVISNAMAAYSSARLLVRGDEPEEKPSADVHOLVP | -----PRMLERFTPKVKX | ----- | 5384 | | |
| FT | ACT_SITE | 1585 | 1585 | BY SIMILARITY. | Db | 583 | CLAENALGOVSCSAWTV | -----HEKASRSKSEVILLVAPSKPTAFIFLOGLSDLKWD | ----- | 636 | | |
| FT | DOMAIN | 1906 | 1914 | POLY-GLU. | Qy | 5385 | GSSITFSVKVEGRPVPTVHMLREABRGVLWIGPDTPOYTVASSAQOHSVLVLDVGRHQ | ----- | ----- | 5444 | | |
| FT | DOMAIN | 868 | 898 | I-1. | Db | 637 | GSQVMTVQVSGNPPPEVIMLHNGNEI | -----QESDFHFEQRTGCHSLWIOEVFPEDT | ----- | 690 | | |
| FT | REPEAT | 896 | 895 | I-2. | Qy | 5445 | GYTCTIASNAGQALCSASLHVS | -----GLPKVEQEKVKALISTFLOG | ----- | 5489 | | |
| FT | REPEAT | 924 | 923 | I-3. | Db | 691 | GYTCEANASAGEVTRQAVLTVOEPHDGTQVWFISKPSRVSTASLGQSVLLSCALAGDPFP | ----- | ----- | 750 | | |
| FT | REPEAT | 952 | 951 | I-4. | Qy | 5490 | TTQAI | -----SAQLETFASADLGQKKEEPLAAKALGHLSLAEVGTBEFLQKLTQSITEM | ----- | 5546 | | |
| FT | DOMAIN | 999 | 998 | I-5 (INCOMPLETE). | Db | 751 | TVHWRDCKALCKDTGHFEVLQNEDEVTLVLKYPQVHAGQVEILLKXRVGCECQVSLM | ----- | ----- | 810 | | |
| FT | DOMAIN | 999 | 1063 | 6 X 12 AA APPROXIMATE TANDEM REPEATS. | Qy | 5547 | VSAKITOAKLOVRGG | -----DSDED | -----SKTPASPRHGRSRPSSSIQ-ESSSESDG | 5595 | | |
| FT | REPEAT | 999 | 1003 | II-1 (INCOMPLETE). | Db | 811 | LQNSARA | -----LPRGREPASCEDLCGGVGADGGSDRYGSLRFGWPARGGOWLEEDGE | ----- | 867 | | |
| FT | REPEAT | 1004 | 1015 | II-2. | Qy | 5596 | DARGEIFDIYVTADYPLGAEQDAITLREGQYVEVLDAAHPLRWLVRTKPTKSSPSRQ | ----- | ----- | 5655 | | |
| FT | REPEAT | 1016 | 1027 | II-3. | Db | 868 | DVRG-VLKRRETVRQH | -----TEBAIRQOE | -----VEQLDFRDLGLKKVSTK | ----- | 908 | |
| FT | REPEAT | 1028 | 1039 | II-4. | Qy | 5656 | WVSPAYLDRRLKLSPEWGAAPPEPGEAVSEDEYKARLSVIOELLSSEQAFAVEELOPL | ----- | ----- | 5715 | | |
| FT | REPEAT | 1040 | 1051 | II-5. | Db | 909 | -----TLSED | -----DLKEIPABQM | -----DFRANLQVQVPKTVEE | ----- | 941 | |
| FT | REPEAT | 1052 | 1063 | II-6. | Qy | 5716 | QSHHLQLERCPHVPVIAVAGOKAVIFRNVRDGRFHSFLQELQCCDTDDVAMCFIKQ | ----- | ----- | 5775 | | |
| FT | VARSPLIC | 437 | 506 | VSGIPKPEVAMFLEGTVPVRRQEGSIEVYDAGSHYLCLLKA | Db | 942 | -----ERKVHSP | -----QQVDFRS | ----- | A 977 | | |
| FT | FT | FT | FT | RTSDGTYSCTASNAQGVCSWTLQVER -> G (in isoform 2 and isoform 3B). | Qy | 5776 | AAFOYLEFLVGRVOASVVVSTAIQEFYKKYABEALLAGDPSPPPPPHLYLEQPVPR | ----- | ----- | 5835 | | |
| FT | FT | FT | FT | DEVEUSD -> MKWRCT (in isoform 3A, isoform 3B and isoform 4). | Db | 956 | -----VLAKGTSKTPVE | -----KVPRPKP | ----- | 1027 | | |
| FT | FT | FT | FT | /FTid=VSP_004791. | Qy | 5836 | VQRYOAILKELIRKARNRQNCALLEQAYAVVSALPQAEKLVHVSIMENYP | -----GTLEALG | ----- | 1098 | | |
| FT | FT | FT | FT | /FTid=VSP_004792. | Db | 978 | TPDFRVLGGKKLPAPENGSSAETLNKAVESKP | -----LSNQPSGLPKPVG | ----- | 1156 | | |
| FT | FT | FT | FT | /FTid=VSP_004793. | Qy | 5954 | EPTRQGHFIWEGAPGARMKPNHRVFLFRNHLVICKPRDSRTDTVSVPFRNMKLS | ----- | ----- | 1216 | | |
| Query Match | | | | 2.5%; Score 1023; DB 1; Length 1914; | | | | | | | | |
| Best Local Similarity | | | | 21.1%; Pred. No. 2.2e-27; | | | | | | | | |
| Matches | | | | 492; Conservative 293; Mismatches 815; Indels 736; Gaps 73; | | | | | | | | |
| Qy | 4609 | HLPTQVRLAEPK | -----PVPPQPS | -----APESRQVAAGEDVSLELVVA-EAGEVIVHKG | 4660 | 5895 | EPTRQGHFIWEGAPGARMKPNHRVFLFRNHLVICKPRDSRTDTVSVPFRNMKLS | ----- | ----- | 5954 | | |
| Db | 11 | HISXLSLVDSRVS | -----NPLTEAPAFILPPRLNLCIKEGATAKAFGRVGRGYPEPQVTHRNG | 70 | 1028 | 5955 | SIDLNDQEGDDRAFEVQWQREDSVRKYLQARTAILKSSVWKEICIGIQRLALPVRPP | ----- | ----- | 6014 | | |
| Qy | 4661 | ERIQGGRFVVSQGRQM-LVIKGTAEQCEYHGLAOGSICPAAATFOVALSPASVD | 4719 | 112 | 1064 | KEELKQVNDVNCRKHAGTTDNEKSESQ | -----TA | ----- | ----- | 1098 | | |
| Db | 71 | QPIISGGRFLDCCGIRGTFSLVITHAVEEDRGKYTCATNGS | ----- | 112 | 6015 | DFEELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDPHILIEDPGS-CALI | ----- | ----- | 6073 | | | |
| Qy | 4720 | EAPQSLPPEAAQEGDLHLLWEALARKRRMSRPTLDSISELPEEDGRSQRLPQEAEEVA | 4779 | 118 | 1099 | AFKQLQDVHVAEGKLLQCVSSDPPATIIWLNKTLTKTKTILLSQ | ----- | ----- | EGSLCSVS | 1156 | | |
| Db | 113 | ----- | -----GARQVT | 118 | 6074 | LDSITGVDSQGYMCFAASAAGNCSITGKILVQVPPFRVN | ----- | ----- | KVR | ----- | ASPPVEGE | 6123 |
| Qy | 4780 | PDLS-EGYSTADELARTGDADLSHTSDDESRACTPSLVTLKAGRGPTSPKASVGP | 4838 | 137 | 1157 | IEKALPEDRGLKCYAKVANDAGQAECSQVTVDDAPASENTKAPEMKSRPKSLPPVLGT | ----- | ----- | 1216 | | | |
| Db | 119 | VELTVEG | -----SFALQGLQPVVS | ----- | 137 | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| Qy | 4839 | AAPSVKQQQOEPLAAVRPPLGDLSTKDLGDPSPMDKAAVKIQAFKGYKVRKEMKQCEP | 4898 | 161 | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| Db | 138 | ----- | -----KTLGD-RFSASAVETRESI | ----- | 161 | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| Qy | 4899 | MFSHTFGDTEAQVGDALRLCIVASKADVRARWLKDGVELTDGRHHHIDQLGDTCSLLI | 4958 | 220 | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| Db | 162 | KFATKLGVRVVKQMGGRFSCKITGRPQVQVTLKGNVPLQPSARVSSE-KNGMQVLEI | 220 | 5018 | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| Qy | 4959 | AGLDRADAGCYTCOVNSKFCQVTHSACVVVSGSEASESSGGLDDAFRABRLHLRF | 5018 | 271 | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| Db | 221 | HGVNQDDVGYTCLVNGSGKASAELSIQGLDSANRS | ----- | 5057 | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| Qy | 5019 | RTKSPAEVSDEELFLSADGPA | ----- | 5057 | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| Db | 272 | R-KEVTNVIKESKJLSLEAAKSKNCSSPQGGSPPPWAANSQPPRESKLESCKDSR | 330 | 5112 | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| Qy | 5058 | ICIRFEALTARQAVT | ----- | 5112 | ----- | ----- | ----- | ----- | ----- | ----- | ----- | ----- |

Query Match
 Best Local Similarity 2.5%; Score 1023; DB 1; Length 1914;
 Matches 492; Conservative 293; Mismatches 815; Indels 736; Gaps 73;

| | | | | |
|----|-----------------------|--------|--------------|---|
| FT | D0VAIN | 1700 | 1763 | CALMODULIN-BINDING. |
| FT | D0VAIN | 1896 | 1906 | POLY-GLU. |
| FT | MOD_RES | 1748 | 1748 | PHOSPHORYLATION. |
| FT | MOD_RES | 1762 | 1762 | PHOSPHORYLATION. |
| FT | CONFLICT | 1439 | 1439 | R -> Q (IN REF. 4). |
| SQ | SEQUENCE | 1906 | AA; 210445 | MM; AD7D8A3B69EE3363 |
| | | | | CR64; |
| | Query Match | | 2.5%; | Score 1012; DB 1; Length 1906; |
| | Best Local Similarity | | 19.7%; | Pred. No. 5.3e-27; |
| | Matches | 488; | Conservative | 318; Mismatches 860; Indels 816; Gaps 76; |
| QY | 4437 | PLKNA | VRAGAQA | RFTCTLSEAVPGEASVYINGAAVQPDSDSWTVTADGSHQ---ALLL 4493 |
| DB | 33 | PPNIR | VQLGATAR | FEGKV--RGYPPEQITWYRNGHPL-PEGDHYV--DHSIRGIFSLVI 88 |
| QY | 4494 | RSAQPH | HAGBVT | IFACRDVASAFLTVGLPDPPEDAEVAHSHSHTVLSWAAPMSDGGG 4553 |
| DB | 89 | KGVQEG | DSGKVT | CEAANDGGVRQTV-----ELTVEGSLKXISLPSSAKTPGGR 139 |
| QY | 4554 | LCGYR | VEVBEGAT | QWRLCHELVPGPECVVDGLAPGETYFRVAAPVPGAGEFVHLPT 4613 |
| DB | 139 | LSVPVP | VEHRPS | IWGE--SPPKFATKENRVV--VREGQTGRFSCKITG----- 181 |
| QY | 4614 | VLRAE | PKPVPP | OPSPAPESRQVAGEDVSLLEVVAEAGEVIMHKGMERIQPGREPVVS 4673 |
| DB | 182 | | | ---RPQP-----QVTTKGDHLLQONERNMFE 206 |
| QY | 4674 | QGRQO | MLVIKG | FTAEDQGEYHCLAGQSGICPAATFQVALSPASVDEAPQ-SLPP---- 4728 |
| DB | 207 | KTGIO | YLEICNV | QLADAGIVTCTVNSAGKASVASELTVQGPDKDTHAQPLCVPKPKTT 266 |
| QY | 4729 | ---EAA | QEGDLH | LLWEALARKRMSREPTLDSISELPEDGQRSORLPQABEAVAPDLSEG 4785 |
| DB | 267 | LATKA | LENSDF | -----KOATNSGIAK--ELKSTSTELMVTETDRLSAKKET 310 |
| QY | 4786 | YSTADE | L---ART | GDADLSHTSDDESRACT--PSLVTYLKKAGRPTSPSLAKVGAAP 4841 |
| DB | 311 | FYTSR | EAKDKG | KQONQANAVLOESR-CTKGPOVLQ-----KTSSTITLQAVRAQPF 363 |
| QY | 4842 | SVKPO | ---QOE | PLAAPPGLDSTKOLGDPDM--DKAAVKIQAAFKGYKVRKEMKQE 4896 |
| DB | 364 | KAEPOT | TFIROA | EDRKRTVQPLMTTTTQE--NPSLTGQVSPRSRETNENRAGVRKSYKEE 421 |
| QY | 4897 | ---GPF | SFTFGT | EAQVGAALCECVASKADVRAWRLKDGVELTDGRHHHDLQ 4950 |
| DB | 422 | REPLG | TPPFES | RPSQLEASGEQEKFKSKVSGKXPDVENFKGVPYIKTGEGIQIYE-E 480 |
| QY | 4951 | DGTC | SLIAGL | DRADAGCVTCQVSNKFGQVTHSACVWVSGSEASESSGGELDDAFRA 5010 |
| DB | 481 | DGTHC | LWLKKA | CLGDSGSYCAAFNPRGQTST----- 513 |
| QY | 5011 | ARRLH | ELFTK | SPAEVSDEELFLSADGEPAEPEADWQTYREDHFICIRFEALTEARQ 5070 |
| DB | 514 | | | -----W-----LUTVRKP 521 |
| QY | 5071 | AVTRF | OEPMFAT | LIGVIEIKLVEGQPRRVMCISKETPAVPVPPEPLSLTSDAAPVFLT 5130 |
| DB | 522 | KV | | -----EEVAPCFSS 532 |
| QY | 5131 | ELQNE | VQDGPY | SPDCVVTGQPMFSVRWFKDGKLLBEDDDHYNINEDQOGHQLIITAVV 5190 |
| DB | 533 | VLKGT | VSGDQF | VLQCVYGVGPVPEITWL-----LNEQPIQAHSTFEAGVAKLTVDQDAL 588 |
| QY | 5191 | PADM | GVYRCL | AEANSMGVSTKAELRVDLTS TDYDAA DATESSYFSAQVYLSRSEQET 5250 |
| DB | 589 | PEDD | GIYTCL | AEANNAGRASCQAQTV-----XEKSSXKAEGT 626 |
| QY | 5251 | EST-T | DEGO | LPOVBEELRDQLVAPQTRIAKFLQKYGKYPAPRLYWFKDGQPL--TASAH 5307 |
| DB | 627 | QAAK | LKNTFAP | IFLXGLTDLKVMDSQVI-MTVEVSANPCPEIIMHNGKTIQETEDFHF 685 |
| QY | 5308 | RMTG | KKILHT | LEIISVTRDEDSQYAAVYSNANGAAYSARLLVVRGDPDEPEKPSADVHEQ 5367 |

686 EKKGNE--YSLYIQEVPEPDGKYTCBAMNELGETQATLTVQEPDQG----- 732
5368 LVPRMLERFPKVKKSGSITPSVKYVGRPVPTVHMLREAEARGVLMIGDPGTVAS 5427
733 -IQPWFISKRSVTAAGQNVLSICATAGDPFPFVHFKDQGE-----ITPGT-GCEILQ 785
5428 SAOQHSVLVDYGRHQGTCTCIASNAAGALCSALHVSGLPKVBEQKVKREALISTFL 5487
786 NEDIFTLILNRVQSRHAGQYEIQLNRVQGCSCQVSL-----MLRESSASRAEML--- 835
5488 QGTTQALSAOGLFASADLGGQKKEEPLAKAELGHLSLAEVTEFEFLQKLSQITEMV 5547
836 -----RDGRESA-----SGERRD-----GNYGALTGRTSGFKK----- 866
5548 SAKITQAKLQVPGGSDSDSKTPASPRHGRSRPSSSIQSSSESDGDARGEIFDIYV 5607
867 -----SSSETRAEEQEDYRG-VLKERVE 890
5608 TADYLPICAEODAITLREGQVVEVLDAHPLRLVIRTKPKSSPSRQGWSPAYLDRRLK 5667
891 TREH-----TEB-----SLAQ-QBAEQLDPRDILGKKVSTK-----SFSEEDLK 928
5668 LSPWGAAPPEPFGAEVSEDEYKARLSSVIO-ELLSQAFVEELOFLQSHLQHLERC 5726
929 -----EIPAEQK-----DFRANLQKQVKTLSSEERKHPAQOVDFRSLAKGKT 974
5727 PHVPI--AVAGCKAVI--FRNVRDIGRHSFLOEQOCDTDDVAMCFIKNOAFAEQYL 5782
975 KPTPLPEKVPKPAVTDFRSLGAKK----- 1001
5783 EFLVGRVQAESVVVSTAIQBYKYABEALLAGDPSPQPPPLQHYLEQFVERVQYQAL 5842
1002 -----KPPAENGSASTAPN--ARAGSEAQVATNSAPAP----- 1035
5843 LKELIRKARNRQNCALLEQAYAVVSA--LPQRAENKLHVSLMENYPTGLEALGEPH 5900
1036 -KPVVKEEKNDKRC--EHCAVVDGIIIGKKAENK----- 1068
5901 HFIVWGAAPGARMKPNKHNHVFRLPHNLVICKRRDSRTDTSYVFRNMKLSIDLND 5960
1069 -----PAASKP-----TPPPSKGT----- 1082
5961 QVEGDDRAFEVQWQREDSVRKYLQARTAIKSSWKEICIGIQORLALPVMRPDPFEEL 6020
1083 -----VPPRFVNRKVRASPEVE-----GEDAQFTCTIEGAPYQIRWYKDGALLTGN 6152
6021 ADCTAELGETVKLACRVGTGPKVISMWYKDGKAVQVDPHILIBDPGSCALILDSLTGV 6080
1091 QDAKVADGKELVLQCRISDDPPASVSWTLDSKAIK-SSKSIVISQEGTLCGLATIEKWPE 1149
6081 DSGQYMCFAASAAGNCSTLGLILNQ----- 6105
1150 DGGETYKCTAENAGKAEACACKVLVEDTSTKAAPKAEKTKKPTTLPVVLSTESSATV 1209
6106 -----VPPRFVNRKVRASPEVE-----GEDAQFTCTIEGAPYQIRWYKDGALLTGN 6152
1210 KKKPAPKTPKAATPPQITQFPEDKVRAGESVELFAKVGTAPITCTWKFQKQIGENE 1269
6153 KFTLSEPRSGLLVLPAAASKEDGLYECVLNRLGASARASALRIQSPMLQAOEQCHR 6212
1270 YIKIENAEISS--KLTISSTKOEHCQCYTLVVENKLGSRQAVNL----- 1312
6213 EQLVAAVEDTTLERADQEVTSVLKELGPKAPGSPSTGLTGPGRGAPALQBTGSQPP 6272
1313 -----TVVDKED-----PPAGTFCASDIRS-----SSLTSLWVGSS-- 1343
6273 VTGTSEAPVPRVQPLLHGGPEQPEAIARAQEWTPVIRNVEGAAMPAGTGBELLND-V 6331
1344 -----YDGG-----SAVOSYTVET-----WNSV 1361
6332 HSHVRETTQTYTQAITDTHAREPSSMQVITEDVQAOTGGTAQFAEALIEGDPQPSWTY 6391
1362 DNKWTDLTCTSTSNVQLOADREYKFRVRAANVYIGSEPSQSEVVKVE----- 1413

6392 KDSVOLVDSTRLSQQOEGTYSVLVRHVASKDAGV-YTCLAQNTGGQVLCKAELVLQGD 6450
1414 QOEELKEEAEALSDEG-----KETEYVYRTVINT----- 1445
6451 NEPDSEKSHRRKHLHSFVEVKEEIGRGVGFVKRVQHKGNKILCAAKFIPLRSRTAGAY 6510
1446 -----EOKVSDVYNIERLGSRGFGQVFLVEKKTGKVMAGKFFKAYSKEKENI 1495
6511 R-ERDIALSHPLVTGLDQDPETRKTLILILELCSSEELDLRLYRKGV-VTEAEVKVYI 6568
1496 RREISIMNCLHHPKLVQCVDAFEERKANIVMLVEMVSGGELFERIIDEDELFETERECIKYM 1555
6569 QOLVTEGLHYLHSHGVHLHDIKPSNLMVHPAREDIKICDFGPAQNITPAELQFSQYGSPE 6628
1556 ROISEGVEYIHRQGVHLDLKPENIMCVNKTGTSIKLIDFGLARLEASAGSLKVLFGTPE 1615
6629 FVSPBIIQONPVSEASDIWAMGVIYSLSLTCSPPAGESDRATLNLVLEGRVSWSSPMAA 6688
1616 FVAPEVINVEPIGYETDMWSIGVICIYLVSGLSPMGNDNETLANVTSATWDFDEAFD 1675
6689 HISEDAKOFIKATLQAPQAPSAOQCLSHPWFLKSMFAEEAHFINTQKFLLLARSWQ 6748
1676 EISDDAKDFISNLLKDKMKRLNCTQCLQHPWLQDKTKNMEAKKLSDRMKKYMARRKWQ 1735
6749 RLSMYKSLVNRSPPELLRGPDPSPSLGVAHLCEDTGGSSSSSSSDNELAPARAKS 6808
1736 KTHAVRAI-----GRSSWAMISGMSORKASGSS----- 1765
6809 LPPSPVTHSPLLHPRGFLRPSA 6830
1766 -PTSPINADKVENEDAFLEEVA 1786

RESULT 5

TRIO HUMAN
ID TRIO HUMAN STANDARD; PRT; 3038 AA.
AC Q75952; Q13458;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Triple functional domain protein (TPPRF interacting protein).
GN TRIO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Fibroblast;
RX MEDLINE=96224308; PubMed=8643598;
RA Debant A., Serra-Pages C., Seipel K., O'Brien S., Tang M., Park S.-H.,
RA Streuli M.,
RT "The multidomain protein Trio binds the LAR transmembrane tyrosine
RT phosphatase, contains a protein kinase domain, and has separate rac-
RT specific and rho-specific guanine nucleotide exchange factor
RT domains."
RL Proc. Natl. Acad. Sci. U.S.A. 93:5466-5471(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Streuli M.,
RL Submitted (SSP-1998) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.
RX MEDLINE=99005194; PubMed=9790533;
RA Liu X., Wang H., Eberstadt M., Schmichel A., Olejniczak E.T.,
RA Meadows R.P., Schervantz J.M., Janowick D.A., Harlan J.E.,
RA Harris E.A.S., Staunton D.E., Resik S.W.,
RT "NMR structure and mutagenesis of the N-terminal Dbl homology domain
RT of the nucleotide exchange factor Trio."
RL Cell 95:269-277(1998).
RN [4]
RP CHARACTERIZATION.

Db 1758 DSDSAA*PQDETVEERGRNEGLSSGTLKSSSSGMSQCGEEGE-----EGADAVPLP 1811
Qy 5615 ---GAEQDAITLREQQYVEVLDAHAHLRLWLTPTKTS-SPSRQGWYSP-----AY 5661
Db 1812 PPMALQHSLLQPSQ-----DDKASSELLVR--PTSSETPSAELVAELVKSMAL 1864
Qy 5662 LDRRLKLSPEWGAABEF-PGE-----AVSEDEYKARLSS-----VIQELLSSEQA 5707
Db 1865 EDRPSSLLVDQDSSSPSPNSDNLSSSSPIDEMBEERKSSSLKRHHYVLQELVETERD 1924
Qy 5708 FVEELQFLOSHLQHLERCPHVPITAVAGXAVIERNVRDGRPHSS-FLQELQOCDTD-D 5765
Db 1925 YVRDLGVVEGYMA-LMKEDGVDDMGKQKIVGNTHQIYDHRDFLGELEKCLDEPE 1983
Qy 5766 DVAMCFIKQAAFOYLEFLVGRVQAEVVVYSTAIOEFYKYABEALLAGDPSQPPPPPL 5825
Db 1984 KLGLSLFVCHERRLHMYIAYCQNKPESEH-IVSEVIDTFEELKQK-----L 2028
Qy 5826 QHYLE-----QPVERVQRYQALLKELIRNARNONCALLQKQAYAVVSALPQAEKNKLH 5879
Db 2029 GHRQLQTLTLKPVQRINKYQLLKQFLKYSKASLDTSELAVERAVENCIVPRCNDMMN 2088
Qy 5880 VSLMENYPTGLEALGEPITROGHFTVWEGAPGARMPWKGNRHVFLFNHLVICKPRDSDR 5939
Db 2089 VGRLOQPGDKIVAQKULLQDTFLVTDQDAGL-LP-RCRERIRFLFEQIVIFSEPLDKK 2146
Qy 5940 TDTV-SYVFRNMKLSIDLNDQVEGDDRAPEVQWEREDSVKYLQARTAIKSSWKE 5998
Db 2147 GFSMPGFLFKNSIKVSCLEENVENDPCKFALTSRTGDVETTELHSSSPSVRQTWIE 2206
Qy 5999 ICGI---QORLALPWPRPDEE-----LA 6021
Db 2207 INQILENORFNALTSFVQRHNSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2266
Qy 6022 DCTAELGVTKLACRVGTGTPKPVISWYKQKAVQVDPHH--ILIEDPDGSCALILSLTG 6079
Db 2267 SCG---GAPSTSRSPRIPQPV-----RHHPPVLVSSAASSQA-EADKMSG 2309
Qy 6080 VDSGOYMCFAASAAGNCSTLKIIVQVPPRVNKV-----RASPVEGEDAQCTCTIEG 6133
Db 2310 T-----STPGSL--PPGAAPGSPAPRRPPGADAESE-----REA 2347
Qy 6134 APYQIRWYKDGALLTGNKQTLSEPRSGLLVLVIRAA-----SKEDLGLY 6180
Db 2348 EPIP-----KMKVLESPEKG-----AANASGSSPDAPAKDARASLGLT 2385
Qy 6181 ECELVNLGSRASAELEIQSPMLQAOEQCHREQLVAAVEDTTLERADQEVTVLKLKG 6240
Db 2386 P-----LGKPRAGAASPINSPLSAPVSLGKEPPPS---SPLQKGSFWSST----- 2430
Qy 6241 PKAPGPGDGLTGPQ-----PCPRGA-----PALQETGSGPP- 6272
Db 2431 PASPASRPGSTFFGSDSLORQTPRHAAPGKDTDRMSTCSSASEQSVQSTQNGSESS 2490
Qy 6273 -----VT-----GTSEAPAVPPRV-- 6287
Db 2491 SSNISTMLVTHDYTAVKDEENVYQVEVQVILASNQNMFLVFRATDQCPAAGWIPGF 2550
Qy 6288 -----QPLHEGPEQEPARAOEWTPVEM-----EGAAWPGAGTGL- 6331
Db 2551 VLHTSAVIVENPD---GTLKKSSTWHTALRKKSEKOKDKRGKLENGYKSRGL 2607
Qy 6332 HSHVVRETTQRTYTYCAIDHTARTPPSQVITDVQAGTGAQFAEIIIEGDPQPSVTWY 6391
Db 2608 SNKVSVKLNNPYIDV-----PPEFVIPLEVTCTGTETVVLRCVCRGRPRASITWK 2660
Qy 6392 KDSVOLVD-----STRLSQOQEGITYSLVLRHVASKDAGVYTCIAQNTGGQVLCABLV 6446
Db 2661 GPEHNTLNDGHYSISYDLGEAT---LKIYGVTTEDDGIYTCIAVNDMSGASSASLRV 2717
Qy 6447 LGGDNEPDSQKSHRKLHSEFVEKKEICRGVFGKRVQKHGKNIKILCAAFIPLRGRTR 6506
Db 2718 LGPG---MDGIMVTKDNFDSYSEVAELGRGFSVVKCDQKGYKRAVATKFNKLMKR 2775

Qy 6507 AOAYRERDILAALSHPLVTGLLDQFETKRLTLILIELCSSEBLLDLRYKGVVTEAEKV 6566
Db 2776 DQVTHELGILQSLQPELLVGLDTPPTPSYILVLEMDQGRLLDCVVRWGSITGKIRA 2835
Qy 6567 YTCQOLVEGHYHSHGVHLHDIKPSNIMWHP-AREDIKICDFGAQNTITPBLQFSQYG 6625
Db 2836 HLGEVLEAVRYLHNCRIAHLDLKPENILVDESAPKTXIKLADFGDAVQNLNTTYIHQLLG 2895
Qy 6626 SPEFVSEPIIQNPVSEASDIWAMGVISYLSLTCSPPFAGESDRATLLNVLGRVSWSSP 6685
Db 2896 NEFEAPAEIILGNFVSLTSDTWSGVLTIVLSGVSPFLDSDSVETCLNCRDPSFPDD 2955
Qy 6686 MAHSEDAKDFIKATLQAPQAPSAOCLSHPMFLKSMPEABEAHFINTKOLKELLARS 6745
Db 2956 YFKGVSKAKFEVCFILQEDPAKPSAALQEQW-LOAGNGRSTGVLTSLTSLTFIERR 3014
Qy 6746 RWQ---RSLMSYKSLIWMESIPEL 6766
Db 3015 KQNDVRPIRSINKFLOSLLPRV 3038

RESULT 6

PGEM MOUSE ID PGEM MOUSE STANDARD; PRT; 3707 AA.
AC Q05793; 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate
DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=92078153; PubMed=1744087;
RA Noonan D.M., Fullie A., Valente P., Cai S., Horigan E., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "The complete sequence of perlecan, a basement membrane heparan
RT sulfate proteoglycan, reveals extensive similarity with laminin A
RT chain, low density lipoprotein-receptor, and the neural cell adhesion
RT molecule";
RL J. Biol. Chem. 266:22939-22947(1991).
RN [2]
RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89034110; PubMed=2972708;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "Identification of cDNA clones encoding different domains of the
RT basement membrane heparan sulfate proteoglycan";
RL J. Biol. Chem. 263:16379-16387(1988).
CC -I- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -I- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -I- SUBCELLULAR LOCATION: Extracellular.
CC -I- TISSUE SPECIFICITY: Found in the basement membranes.
CC -I- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -I- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -I- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -I- SIMILARITY: Contains 3 laminin IV domains.
CC -I- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC -I- SIMILARITY: Contains 3 laminin G-like domains.
CC -I- SIMILARITY: Contains 1 EGF-like domain.

FT DISULFID 1886 1932 BY SIMILARITY.
FT DISULFID 1976 2021 BY SIMILARITY.
FT DISULFID 2073 2118 BY SIMILARITY.
FT DISULFID 2170 2215 BY SIMILARITY.
FT DISULFID 2268 2313 BY SIMILARITY.

Query Match 1.9%; Score 767; DB 1; Length 3707;
Best Local Similarity 20.8%; Pred. No. 2.2e-18;
Matches 870; Conservative 464; Mismatches 1543; Indels 1308; Gaps 218;

QY 1193 POTEVWYKDGKLSKSSKVRMEVKGCTRRLLVVOQVGKADAGEYSCBAGQORVSFOLHIT 1252
DB 98 FOLEDASAKREFVESEAVVEKLE-----PEYRKIPGQIVS----- 133

QY 1253 EPKAVFAKE-----OLVHNEVTEAGASATLSCEVAQAQTEVWYKDGKLSKSSKVRIEA 1308
DB 134 ---VVFIEKELGOWFVELDVGSEGNAGSQIEV--LHTVVSSGSIOPYVTSPNGFKFR 188

QY 1309 AGCMRQ-----LVVQOAGADAGEYTC--AGGORLSFHLVDSEPKAVFAKEQLAH 1357
DB 189 LGTVPPRPVCTETEFACHSYNECVALEYRCRRDPDCRDMDELNCEPVPPELSSSTPA- 247

QY 1358 RKVQAEAGAIATLSCEVAQAQTEVWYKDGKLSKSSKVRMEAVGCTRRLLVVOQAQOAT 1417
DB 248 ---VGKVSPLPLMPEAATPPPPVTHGPQFLPLSPVGPSCAG-----P-287

QY 1418 GEYSCEAGQRLSPSLDVAEPKVFVFAKEOPVHREVOAQAGASTTSLCEVAQAQTE----- 1472
DB 288 QEASCHSH-----CIPFDVLCQOECDRGSDELGCASPPPCPEPNEFACENGHCA 338

QY 1473 -VMYKDGKLSKSSKVRMEAVGCTRRLLVVOQAGQ-ADAGYSCEAGSQRLSFHLVAE- 1529
DB 339 LKLWRCDG---DFDCEDRTDEANCS---VKQPEGVCGPTHFQCVSNRCIPASFHCDEE 391

QY 1530 ---PKAVFAKEQASREVQAEAGTSATLSCEVAQAQTE-VWYKDGKKL 1574
DB 392 SDCPRDSDFGCMPPQV---TPFQSIQASRGQTVTTCVATGVPTPIINMLNWHGI 447

QY 1575 SSSKVRMEAVGCTRRLLVVOBAGADAGEYSCAKADQD-LSFHLHVAEPKVFVFAKEQPAH 1633
DB 448 PAHPRVWTSSEGRGLIIRDVKEADQAGYTCAMNSRGMVFGIPDGVLLELPVQGPCCP 507

QY 1634 REVQAEAGASATLSC-----EVQAQTEVWYKDGKLS----- 1667
DB 508 GHFYLEDSASC-LPCFCFGVTVNVCQSLR---FRDQIRLSFDQPNDFKGVNVTWPSQGV 563

QY 1668 ---SSSKVRV-----EAVGCTRRLLVVOQA-----GQADA-----GEYSCE 1699
DB 564 PLSSTQLOIDPALQEFOLVLSRRFLVHDAFWALPKQFLGNKYDSYGFILYKVRVELA 623

QY 1700 AG-----GQRLSFRILH-----VALEPQISE-----RPCR-EP 1728
DB 624 RGMLEPVQKPDVILVAGYRLHSRGHTPTHPTGLNQVQLSEBHWVHESGRPVQRAEML 683

QY 1729 VVKEHEDJILTATLTPSAATVWLDKGVIEIRRSKRHETASQGDTHLTLVHGAQVLDLAI 1788
DB 684 QALASLEAVLLQTVNTKQASV-GLSDIV-----MDTTVTHT-TLHGRA---HSV 728

QY 1789 YSCR--VGAEGQDFPVQVEEAAKFCRL-----LEPV----- 1818
DB 729 BEECRCPIYSG---LSCSCDAHFTVPVGGPYLGTCSGNCNGHASSCDPVYGHCLNCQ 784

QY 1819 -----C-----GELGTVILACELSP-----ACAE 1838
DB 785 HNTEGPQCDCKKPGFPGDATKATATACAPCPVIDASRRFSDTCFLTDGQATCDACAP 844

QY 1839 VVWRCGNTQPRVGRKFMVAEG-----PV-----RSLTVLGLRAEDAGEY-----VCESR 1883
DB 845 -----GYT-----GRCESCAPYEGNPIQPGKCRPTTQIEIVRCDESGSLTSETCRCK 895

QY 1884 DDHTSAQITVSVPRVVKPMGSLSTVVAEEGEANTQCV---VSPSDVAVVWFRD---GAL 1937
DB 896 NN-----VYGRLCNECSGDSFHLKQNPDGCLKFCMGVSRQCSSSSMSRAQVLGAS 947

QY 1938 LOPSEKFAISQSGASHSLTISDLVLEDA-GQIT-----VBA 1972
DB 948 EOPSQ-FSLSNAAGTH--TTSEGVSPAPGELGFSFPHLLSEPYFWSLPASFRGDKVTS 1004

QY 1973 EGASSSAALRVREAPVLFKKKLEBPOTVERSSVTLLEVELTR-----PWPELFW 2020
DB 1005 YGGELEFVWQPRSSAPLHRQPLVVLOGNNVLEHSHASRDPSPQPSNFVFPQEQAN 1064

QY 2021 TRNATALAGKXVEIHASGARHLRLVHNVGFADRGFFGCETPDDTKOAKLTVMEMQVRLV 2080
DB 1065 QRPDQGFATREHLLMALAGIDALLI-----QASYT----- 1094

QY 2081 RGLQAVEAREEQGTATMEVQLSHADVDGSMTRDGLRFOOQPTCHLAVRGMHTLTLGLRP 2140
DB 1095 ---QQPAESRLSG-ISMVA-----VPENTGQASAREVEOCTCPVGRGSCQDCTGYTR 1146

QY 2141 EDSGLMVFKAEGVHTSARLVVLTPLVFSRPLQDVVTEKEKVTLECELSRNVDRVWLK 2200
DB 1147 VPSGLVGTCCRNCHGHSETCEPETGACQCOH--HTE---GASCEQCPGY----- 1194

QY 2201 DGVLELAGKTMALAAQGA---CRSLTIYRCEADQGVYVC-----DAHDAQSSASVKVQGR 2253
DB 1195 ---YGDQORGTPOCQPCPCYGAPAGQAHAHCTFLDTDGHPTCDSCSPGHSGR 1244

QY 2254 TVTLI-----YRRVL---AEDAGEIQFVAENA-ESRAQLRVKLELV 2290
DB 1245 HCERCAPGYGNPSQPCQCHRDQVPEVLGCGDPHGSISSQDAAQGCQCKAQVEGRSC 1304

QY 2291 TLVRPLRDKIAMEKHGVLEC-----QVRSASAQVRW-----FKG-----S 2326

DB 1305 SHCRPHHFLSASNPBGCLPCFCMGVTQCCASSSRQLISTHFAPGDFQGFALVNPORN 1364

QY 2327 QELQCPKVELSDGLYKLIISD-VHAEDETY---TCDAGDVKTSQFFVEEOSITI 2381
DB 1365 SQTGTGFTVEPHDG--ARLSFNFAHLQESFYWQLEPIYQGD-KVAYAGKLRVLSY 1421

QY 2382 VRGLQDVTVMEPAPAFECETSIPIVRPKWLLGKTVLQAGNVGLGEQGTVHRLMLRRT 2441
DB 1422 TAGPQGSPLLDPD-----IQITGN-----NIMLVAS 1447

QY 2442 CSTMTGVHFTVCKSRSSARLVVDIPVVLTRPLEPKTGRE-----LQSVLSCDFR 2493
DB 1448 QPALQGP-----ERRSYEILFRE--EFWRPPDQGPATREHLLMALADLLELVNATFS 1498

QY 2494 PAFKAVQWYKDDTPIPSSEKFKMSLEG-----QMAELRILRMP----- 2532

DB 1499 SVPERA-----ASISAVSLEGAQPGSPSSGPRALEVEECRPPGYVGLSCQDCAP 1546

QY 2533 ---ADAGVY-----RCQASAHSSSTEVTVAREVTVGPI 2564

DB 1547 GYTRTSGLYLOQCELCNCGHSDLCHPETGACSRQHNTAG-----EFCELCATGY 1599

QY 2565 QAAEA-TEEGWASFSCSELSHEDBEVWSLNGM-----PLYNDSFHEISHKGRH 2612

DB 1600 GDATAGTPEDCQPCACPLTNPNMFSRTCESLGAGYRCTACEPGYTGQYCEQCAPG--- 1656

QY 2613 TLVLKSIQADAGIVRASSLKYSTSARLEVRKVP---VVLKALDDLSAEEGTALQCE 2669
DB 1657 ---YEGDPN-VQGGRCQPLTKESLEVOIHPSRSV-----POGGPHSLRCQ 1698

QY 2670 VS-DPEAHVW-RKDGQVLGSDKDYDFLHTAGTRGLVH--DVSPEDAGLYTCH---VG 2721
DB 1699 VSGSPHYFYWREDGRPLPSSAQORH-----QSELHFPSPVQSDAGVYICTENLIH 1752

QY 2722 SEETRARVRVHD-----LHVGITKTKMEVLEGSCSEFCVLSHESASDPAMWT--VGG 2774
DB 1753 TNSRAELLVAEAPSPIWTVVEEQ-RSQSVRPGADVTFICTAKSPAYTLVWTRLHG 1811

QY 2775 KTVGSSSRQATQGRKYLIVREAPSDAGEVWFSVRL-----TSKASLIYR-----BR 2825
DB 1812 KLPFSRAMDFNG-----ILTRNVQPSDAGTYVCTGSGNFMAMDOGTATLHVQVSGTSTA 1864

QY 2826 PAATIKPLEDOWAPGEDVELCELRSAGTP-VHML--KDKAIRKSOKYDVVCEGTMM 2882
 Db 1865 PVASIH-PQUTVPGQAEFRCSATGNPTMLEWIGPSGQIPAKAQIHN-----GI 1916
 QY 2883 LVIRGASLKADAGEYTCVEASKSTASLHVEE-----RANCFTTELTNLQVEEK 2931
 Db 1917 LRLPAIBSDQGYLCR---ALSSAGQHVARAMLQVHGSGPRVQVSPE---RTQVHEGR 1970
 QY 2932 TAVTCKTEH-PAATVTRK--GLLEIRASQKOPSOEGTLTLTTSALEKASDIYTC 2988
 Db 1971 TVRLCYAAGVPSASITWRKEGSLPFR---HOAHGS-----RLRLHMSVADSGEYVC- 2021
 QY 2989 IQQAQSAQLLVQRRVHIIEDLEVDVQEGSSATFCRISpanyEPVHVFELDKTPLHAN 3048
 Db 2022 -----RAN-----NNIDAQETS-----IMISVSPSTNSP-----PAPAS 2050
 QY 3049 ELNEIDAQPGYHVLTURLQALKDSGTIYFAGQORASALRTEKPSVFSRBLTATIT 3108
 Db 2051 -----PAPIRIESSSSRVA 2064
 QY 3109 EGEDTLVCESTCIDP-----MCWTGDKTGRSARCQLSHEGHAQLLITGATLQDS 3162
 Db 2065 EGQTLDL-----NCVPGHAHAQVTHKRGSL-----PTHQTHGSRURLTQVSADS 2113
 QY 3163 GRYKC-----EAGGACSSIV-----RVH-----ARPRFQEAALKOLEVLEGGATLRVC 3207
 Db 2114 GEYVCVLSGPLEASVLVSITPAANVHIPGVVPPIRIETS--SSRVAEGQTLDLSCV 2171
 QY 3208 LSSVA-APVWCYGNVLRPGDKYSLRQGMLELVNLRPODSGEYSC-----SEGDQT 3262
 Db 2172 VPGQAHQVTHKRGSLPAGHQV-----HGMURL--NRVSPADSGEYSCVQTSSGTL 2225
 QY 3263 TSALTITVA-----LPAQFIG-----KLRNKEATEGATATLRCEL--SKTAPVEMWKGSE 3310
 Db 2226 ASVLVTEASRSPPIAPGLAQPVIYESSSHLTGQTVDLKCVVPGQAHAQVTHKRG 2285
 QY 3311 TLRGDRYCLQDQAMCELQIRGLAMVDAABYSC-VCG-----BERTSASLTIRPMAHFI 3365
 Db 2286 SLP-----ARHOTHGSLLELYQLSPADSGEYVCQVAGSHPEHAFKLTIV-PSQNS 2338
 QY 3366 GRLRHOE-SIB-----GATATLRCELSKAA--PVEWKGRESLRDGRSLRQDGA 3413
 Db 2339 FRLRSPVISIEPPSTVQOGDASFCLIHGAMPKIVENKIRDOELEDNVH--ISPNGS 2396
 QY 3414 VCELOICGLAVD--AGEYSCVCEBERTSATLTWKALPAKTEGLRNEEAVE---GATAM 3468
 Db 2397 IITIVAPGATMEPTACVASNVYMAQSVNLSVHGPP---TVSVLPEGPHVYKMGKDIT 2453
 QY 3469 LWCELSKVPVEMWKGRENLRDGRYTLRQE-----GTRCELOICGLAMADAGEYLC--- 3520
 Db 2454 LEC-----ISSGEPRSSFRWTLGLIPVKLEPRMGLMNSHMLKIASVKPSDAGTYVCOAQ 2509
 QY 3521 -VCQERTSATLTI---RALPARFIEDVKNQE---AREGATAVLQCELSNAP--VEWKG 3572
 Db 2510 NALGTAQXQVELIVDTGTVA PGTPQVQVBESELTLEAGHTATLHCSATGNPPPTIHWK- 2568
 QY 3573 SETLRDGRYSLRODGYKCELOIRGLAMADTGEYSC-----VCQERTSAMLTRALPIKF 3628
 Db 2569 ---LRAPLFWOHRIEGN--TLVIPRVACQDSQGYICNATNSAGHTEATVVLHVESPPYA- 2622
 QY 3629 TEGLRNBEATGATAVLRCELSKAPVEM-WKHETLRDQ---DRSLRDGARGCELQIR 3684
 Db 2623 TIIPHTSAQGNLVOLQCLAHGTPPLTYQW-----SLVGGVLPKAVVRNQ-----LLRLE 2674
 QY 3685 GLVABDAGEYLCMGCKERTSAMLTRAMPKSFTEGLRNEBEATGDTATLM----- 3734
 Db 2675 PTVPEDSGRYCQVSNRNVGSAEFAQVILVQSSSNLPTDSIPGSGSTPTVQVTPQLETRNI 2734
 QY 3735 ---CELSKAAVE-----WRKHETLRDGRSLRQDGRSCLQIRGLAVVADAGEYSCV 3785
 Db 2735 GASVEFHCAVPNERGTHRLWLKEGQLPPG--HSV-QDG---VLRIONLDONCCGYIVCQ 2788
 QY 3786 C-----GOERTSATLTVRALPARFIEDVKN--QEARREGATAVLQD-ELSKAAP-VEWKG 3837

Db 2789 AHGPHGQAQATLQILVQALPSVLI-NVRTSVHVVVHSHVEFECLALGDPQVPTWMSKV 2847
 QY 3838 EFLRGDRYSLRQDGRTRCELOIHGLSVADTGEYSC-----VCQERTSATLTVRAPQPVFR 3893
 Db 2848 GELRPG-----IVQSGT--IIRIAHVELADAGQYRCATNAAGTTQSHVLLVQALPOIST 2901
 QY 3894 EPLQSLQAEESTATLQCELS-EPTATVWSKGGIQLQANGRRBRRLQCGCTAEVLVLOLQ 3952
 Db 2902 PP-BIRVPAGAAVFPMAWGYPITPAITWSKVDGLDPPDSRLNNM-----LMLFSVR 2993
 QY 3953 REDTEYCTC-----GSOATSATLTV-----TAAPVRFRELOHQEVDGEGTAHLCC 4000
 Db 2954 PEDAGTYVCTATNRQGVKAFAYLQVPERVIPYFQTPYSFLPTTKDA----- 3003
 QY 4001 ELSRAGASVEWRKSLQFLPCAKYQMVODGAAALLVRGVEQEDAG----- 4046
 Db 3004 -----YRKFEI-----KITFRPDSADGMLLYNGQKRSPTNLANRQPDFISFGLV 3047
 QY 4047 ---DYTCDTGTOSMASLSVRVRP---KFK--TRLQSLQET---GDIARLCCOLSD 4093
 Db 3048 GRPFRFDAG--SCMA--TIRHPTPLALGOFHTVTLRLSLTQGLIVGNLAPV---N 3098
 QY 4094 AESGAVVQWLKEGVELHAG--PKY-----EMRSQATRELLIH----- 4129
 Db 3099 GTSQKFGQLDLNEELYLGYPDYGAIPKAGLSGFGVCVRELRIQG--BEIVFDVNL 3156
 QY 4130 -----OLEAKDTGEYACVTCGQKTAASLRVTEPEVTVIRGLVDAE- 4169
 Db 3157 THGISHCTCDRPPCQNGGQDQDSESSSYTCVCPAGFTAAAVNIRKCTATPSLWADATC 3216
 QY 4170 VTADEDVFSCEVRAGATGYQWCLQGLPLOSNEVTEVAVRDRGRIHTLRKGVTPEDAGT 4229
 Db 3217 VNRPDGRGYTCR-CHLGRSGVR-CEEGTV-----TTSPMSG 3252
 QY 4230 VSF-----HLGHASSAQLTVRAPEVITILEPQDVLQSEGQDASFQ--CRLSRASSQ-EAR 4282
 Db 3253 SYLALPALTUTHHELRLDV---EFKLEPNGLILFSGKSGPVEDFVSLAMWGCHLEFR 3309
 QY 4283 WALGG-----VPLQANEMNDITVEQ-----GTLHL-----LTLH--- 4311
 Db 3310 YELGSLAVLRSHBELALGRHVRSAERLNKDSLVRDGGPRVLRSPGKSGOGLNLHTLL 3369
 QY 4312 -----KVTLEDAGTVSFHVTCSSEAOQKVTAKNTVVRGLENVVALEGEALFEQLS 4364
 Db 3370 YLGGVEPSVQSPATNMSAHFHGCVGE--VSUQGR-----LDLTYSLGSGVQGYCYS 3422
 QY 4365 OPEVAHAHTWLLDDEPVRTSEN-----ABVWF---FENGLRHLNLLNLRP---QDSRVTF 4414
 Db 3423 SP-----CERQPCRNAGTCMPAGEYEFQCLQDGFKGDLCEHENPCOLHEPC--- 3470
 QY 4415 LAGDMVTSAFITVRGWRLEIIEPLKNAAVRAGAQAARTCTLSEAVPVE-----A 4464
 Db 3471 LMG-----TCRGARCLCPGFSRQCGQ--AGYGVESDWHPEGSGNDAPGYGA 3521
 QY 4465 SWYINGAAVQPDSS-----DMVTADGSHQALLLRSAQPHHA----- 4501
 Db 3522 VFYDNGFLGLPQNSFRSLPVPETIPEVITSTADG-----LLWQGVVREASRKFIS 3577
 QY 4502 -----GEVTFACRDAVASARITVLGLPDPPEAEVAHSHHTVLSWAAPNSDGGGLCG 4566
 Db 3578 LCLQDCHLVFVQLSGEARL-VSG--DPINDGE-----WHRIT-----ALREGQSG-S 3622
 QY 4557 YRVEKEGATGWRLCHLVPGCEVVD-----GLAPGETYFRVAAV--GPVGAEP 4607
 Db 3623 IQVGEDILVTR-----SPGNVAVNTKDIIVYGAPD-----VATLRGAFSSGIT 3669
 QY 4608 VHLPTQVRLAEPKFPVPPQPSAPESRQVAGEDVSLLEVVAEAG 4652
 Db 3670 GCIKNLVLTARPAPPPPP-----LDLQHRQAQ 3699

KWLS BOVIN STANDARD; PRT; 1176 AA.
 ID - KMLS_BOVIN
 AC Q28824;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Myosin light chain kinase, smooth muscle (EC 2.7.1.117) (MLCK)
 DE [Contains: Telokin].
 DE MYLK.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Stomach;
 RX MEDLINE=93203148; PubMed=1284247;
 RA Kobayashi H., Inoue A., Mikawa T., Kuwayama H., Hotta Y., Masaki T.,
 RA Ebashi S.
 RT "Isolation of cDNA for bovine stomach 155 kDa protein exhibiting
 myosin light chain kinase activity.";
 RL J. Biochem. 112:786-791(1992).
 CC -!- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
 SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
 IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
 FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS.
 CC -!- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
 light-chain] phosphate.
 CC -!- SUBUNIT: ISOFORM TELOKIN BINDS CALMODULIN.
 CC -!- ALTERNATIVE PRODUCTS:
 Event=Alternative initiation;
 Comment=At least 3 isoforms, Non-muscle, Smooth-muscle (shown
 here) and Telokin, are produced by alternative initiation.
 Isoform Non-muscle is the longest and telokin is a C-terminal
 section without catalytic activity;
 CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
 CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; S57131; AB25794.1; -
 CC PIR; JN0583; JN0583.
 CC HSP; P56276; 1TLK.
 CC InterPro; IPR008957; FN III-like.
 CC InterPro; IPR003961; FN III.
 CC InterPro; IPR003962; FNIII subd.
 CC InterPro; IPR007110; IG-like.
 CC InterPro; IPR003598; IG-c2.
 CC InterPro; IPR000719; Prot_kinase.
 CC InterPro; IPR008271; Ser_thr_pkin AS.
 CC Pfam; PF00041; fn3; 1.
 CC Pfam; PF00047; ig; 2.
 CC Pfam; PF0069; pkinase; 1.
 CC PRINTS; PR00014; FNTYPEIII.
 CC ProDom; PD000001; Prot_kinase; 1.
 CC SMART; SM00060; FN3; 1.
 CC SMART; SM00408; IGc2; 2.
 CC SMART; SM00220; S_TKc; 1.
 CC PROSITE; PS50835; IG LIKE; 3.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC Transferrase; Serine/threonine-protein kinase; Calmodulin-binding;
 KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;

KW Alternative initiation.
 FT CHAIN 1 1176
 FT CHAIN 1022 1176
 FT INIT MET 1022 1032
 FT DOMAIN 100 291
 FT REPEAT 100 111
 FT REPEAT 112 123
 FT REPEAT 124 135
 FT REPEAT 136 147
 FT REPEAT 148 159
 FT REPEAT 160 171
 FT REPEAT 172 183
 FT REPEAT 184 195
 FT REPEAT 196 207
 FT REPEAT 208 219
 FT REPEAT 220 231
 FT REPEAT 232 243
 FT REPEAT 244 255
 FT REPEAT 256 267
 FT REPEAT 268 279
 FT REPEAT 280 291
 FT DOMAIN 356 444
 FT DOMAIN 498 586
 FT DOMAIN 603 673
 FT DOMAIN 725 980
 FT DOMAIN 972 1035
 FT DOMAIN 1069 1158
 FT NP_BIND 731 739
 FT BINDING 754 754
 FT ACT_SITE 846 846
 FT DOMAIN 1171 1176
 FT SEQUENCE 1176 AA; 128824 MW; F53DC6D4D42D4B97 CRC64;
 Query Match 1.6%; Score 645.5; DB 1; Length 1176;
 Best Local Similarity 23.9%; Pred. No. 78-15;
 Matches 197; Conservative 133; Mismatches 292; Indels 203; Gaps 20;
 QY 6014 POFEEELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDPHHILIEDPGS-CAL 6072
 DB 356 PFEEELQDLHVAEGQKLLQCRVSDPPATITWTLNGTKLTATKTFIVLSQ--EGSLCSV 413
 QY 6073 IDLSLTGVDSQGMCPAASAGNCSTLGLILVQVP--PFPVN-----KVR-----ASPV 6120
 DB 414 SIEKALPEDRLGKVCVAKNSAGQASSSCQTVVDVDPDPTSENAKAPEMKARRPKSLPPV 473
 QY 6121 EGEDAQTCTIGAPY-----POIRWYKDGALLTGN-----X 6153
 DB 474 LGTESDATVKKPAPKPTPKAAMPFOIIFPDQKVRAGESVELFGKVAGTQPICTWVK 533
 QY 6154 FQTLSEPRSGLLV-----LVIRAAKEDLGLVECELNRLIGSARASAEIRIQSPMLQ 6205
 DB 534 FRKQIQDSEHIKVENSEQGSKLTIRAAQEHGCGCYTLLENKLGSRQAQVNL-----595
 QY 6206 AQEQCHREQLVAAVEDTTLERADQEVTSVLKLLGPKAPGPTGDLTGPQCPRGAPALQ 6265
 DB 586 -----TVVDKFPD-----PAGTFCAS 601
 QY 6266 ETGSPPVVTGTSEAPVPRVFPQLLHEGPEOPEAIPAQAQWTVPIRMEGAAMPAGTG 6325
 DB 602 DIRSS-----SLTSLWYSSYDGSVAV 623
 QY 6326 EL-----LWDVHSHVRE--TTQTYTYQATDHTARPPSMQVTTEDVQAQTGGTAQFEAL 6380
 DB 624 QSYSVEMDSVDKTKWELATCRSTSNVQDILLPDREYKFRVRAINIVYGTSEFSQSELT 683
 QY 6381 EGD-POPSVTWYKDSVOLVDSTRLSQOQSGTTYSLVLRHVASKDAGVYTCLAQNTGGQVL 6439
 DB 684 LGKPREEP---KDEVEVSDDEKEFEVD-----YRTVTVT-----717
 QY 6440 CKAEILLVLDGNDPEPDKSEKSHRRKLSHFYVEVEKIEGRGVFGFVKRVQHKGNKILCAKFI 6499

```

Db 718 -----EQKVSDFYDIEERLGSFGQVRLVEKTKGIWAGKFF 756
QY 6500 PLRSTRQAQYR-ERDILAALSHPLVGLLDQFETRTKLILILELCSSEBLLDRLYKGV 6558
Db 757 KAYSAXEKENIRQISINMNCVHLKPLVQCVDPEKANIWMVLEIVSGGELFERIDEF 816
QY 6559 -VTEAEVRYIQQLVEGHLHSHGVHLHLDIKPSNLMVHPAREDIKICDFGAQNTIPA 6617
Db 817 ELTERECIKYWKQISEGVYHKQIVHLDLKPENINCYNKTRIKLIDFGLARLENA 876
QY 6618 ELQFSQSGSPFVSPETIQQNPVSEASDIWAMGVISYLSLTCSPFAGESDRATLLNVLE 6677
Db 877 GSLKVLFGTPEFAVEINTEPIGYATDMWSIGVICYLVSGLSPFMGNDNETLANVTS 936
QY 6678 GRVSWSSPMAHLGEDAKDFIKATLQRAPOARPSAOCILSHPWFLKSPAEAEHAFINTKQ 6737
Db 937 ATNDFDDEAFDEISDDAKDFISNLKMDKMKVRLNCTOCLQHPWLMKDTKMEAKKSKDR 996
QY 6738 LKFLASRWORLSMYKSYLVKRSIPEL-----LRGPPDSP 6774
Db 997 MKYMARCKNKTGNVRAIGRLSSMAMISGLSGRKSSTGSP 1041

RESULT 8
XMLS_RABIT
ID XMLS_RABIT STANDARD; PR7; 1147 AA.
AC P29294; Q28729;
DT 01-DEC-1992 (Rel. 24, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin light chain kinase, smooth muscle (EC 2.7.1.117) (MLCK)
DE [contains: Telokin].
GN MYLK.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_taxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Smooth muscle;
RX MEDLINE=92084694; PubMed=1748666;
RA Gallagher P.J., Herring B.P., Griffis S.A., Stull J.T.;
RT "Molecular characterization of a mammalian smooth muscle myosin light
chain kinase.";
RL J. Biol. Chem. 266:23936-23944(1991).
RN [2]
RP SEQUENCE OF 993-1147 FROM N.A. (TELOKIN).
RX MEDLINE=92084695; PubMed=1748667;
RA Gallagher P.J., Herring B.P.;
RT "The carboxyl terminus of the smooth muscle myosin light chain kinase
is expressed as an independent protein, telokin.";
RL J. Biol. Chem. 266:23945-23952(1991).
CC -!- FUNCTION: Phosphorylates a specific serine in the N-terminus of a
myosin light chain.
CC -!- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
light-chain] phosphate.
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative initiation;
Comment=At least 3 isoforms, Non-muscle, Smooth-muscle (shown
here) and Telokin, are produced by alternative initiation.
Telokin is a C-terminal section with no catalytic activity;
CC -!- DOMAIN: TELOKIN BINDS CALMODULIN.
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC DR W76233; AAA73093.1; -
DR EMBL; M76234; AAA31408.1; -
DR EMBL; M76181; AAA31409.1; -
DR PIR; A41675; A41675.
DR PIR; A59307; A59307.
DR HSSP; P56276; 1TLK.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR000719; Prot Kinase.
DR InterPro; IPR008271; Ser Thr pkin AS.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; Ig; 2.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PRO0014; FNTYPEIII.
DR ProDom; PD000001; Prot Kinase; 1.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00220; S-TKc; 1.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
KW Transferase; Serine/threonine-protein kinase; Calmodulin-binding;
KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
KW Alternative initiation.
FT CHAIN 1 1147
FT CHAIN 993 1147
FT CHAIN 993 993
FT INIT MET 993 993
FT DOMAIN 100 288
FT REPEAT 100 111
FT REPEAT 112 123
FT REPEAT 124 132
FT REPEAT 133 144
FT REPEAT 145 156
FT REPEAT 157 168
FT REPEAT 169 180
FT REPEAT 181 192
FT REPEAT 193 204
FT REPEAT 205 216
FT REPEAT 217 228
FT REPEAT 229 240
FT REPEAT 241 252
FT REPEAT 253 264
FT REPEAT 265 276
FT REPEAT 277 288
FT DOMAIN 329 417
FT DOMAIN 469 557
FT DOMAIN 574 644
FT DOMAIN 696 951
FT DOMAIN 943 1006
FT DOMAIN 1041 1130
FT NP_BIND 702 710
FT BINDING 725 725
FT ACT_SITE 817 817
FT DOMAIN 1138 1147
FT CONFLICT 1114 1114
FT SEQUENCE 1147 AA; 125719 MW; F039E624C6E31024 CRC64;
Query Match 1.5%; Score 622; DB 1; Length 1147;
Best Local Similarity 23.2%; Pred. No. 4.2e-14;
Matches 211; Conservative 136; Mismatches 296; Indels 266; Gaps 25;
QY 6014 PDFEEELADCTAEIGETGKACRVGTGPKPVISWYKDGKAVQVDPHILIEDPDGS-CAL 6072
Db 329 PTFKEKQDVRAVGEKLLCCQVSSSEPPATITWLNKTKTKTKFVILSQ--EGSLCSV 386

```



```

FT CONFLICT 291 P -> L (IN REF. 2).
FT CONFLICT 339 EACHRP -> TDLRGM (IN REF. 2).
FT CONFLICT 659 T -> A (IN REF. 2).
FT CONFLICT 691 A -> T (IN REF. 2).
FT CONFLICT 738 R -> G (IN REF. 2).
FT CONFLICT 923 R -> T (IN REF. 2).
SQ SEQUENCE 1270 AA; 140632 MW; 699947C3C9B58931 CRC64;

Query Match 1.4%; Score 586; DB 1; Length 1270;
Best Local Similarity 20.5%; Pred. No. 7.9e-13;
Matches 329; Conservative 176; Mismatches 562; Indels 534; Gaps 57;

QY 3804 FIEDVQAEQREGATAVLQCELSKA-APVEWRKGSSETLRGDRYSLRQDGTTCRCELOIHL 3862
DB 12 FNKKPRSAEVTAGSAVFEATERSGVMVRWQDGSITANDKYGLAEGKRHTLTVRDA 71
QY 3863 SVADTGEYSCVCGOERTSATLTVRAPQPVFR-----EPLOSQAEEG 3904
DB 72 SPDDQGSYAVIAGSSKVKFDLKVTEPAPPEKAESVAPGAPEEVPAPATELEBSVSSPEG 131
QY 3905 STATLQ-----CELLSEPTATVWS 3923
DB 132 SVSVTQDGAHAGHOGAPDDPTGLFMRPQDGEVTVGGSIVFSARVAGASLLKP-FVVKWF 190
QY 3924 KG-----GLOQANGRRPELQCTAELVLQDLQREDTGEYTCGSG-----ATSA 3970
DB 191 KGWVDLSSKVGQHLQHDSDYASRVYLPFELHITDAQTTSAGGYRCEVSTKDKFDCNF 250
QY 3971 TLTVTAAPVRLRELQHOEYDEGG-----TAHLCELSELRAGASV-----4009
DB 251 NLTV-----HEATGSGDLRSAPRRTSLAGAGRRTSDSHEDAGTDFDSSLK 298
QY 4010 ---EWRKGS-----LQFFCAKYQWV---QDGA---ELLVRGVEQEDAGD 4047
DB 299 KRDSFRDLSKLEAPAEEDVWEILFQAPPSEYERAFQHGVEACHRLKRLKGKQDE---355
QY 4048 YTCDTGHTQSNASLSVVRPRKFKTRLOSLEQ-ETGDIALCCQLSDAESGAVQWLKEG 4106
DB 356 -----KKSTA-----FOKLEPAYQVNNKGKHLRLTVELADPD---AEVKKLKG 396
QY 4107 VELH-AGPKYEMRSGQATRELLIHOLEAKDTGEVACVTCGQKTAASLRVTEPEVTIVRGL 4165
DB 397 QEIQWSGSKYIFESVGAKRYLTISQCSLADDAAYQCVVGGKSTELFVKEPVLITRSL 456
QY 4166 VDAEVTADEVSECEVSRAGATGVQMCLOGLPLOQNEVTEVAV-RDGRHTLRLKGVTP 4224
DB 457 EDQLVMVGQVRVEFECEVSEBGAQ-VKMLKDGVELTREETFKYRFKDKGRKHLIINEATL 515
QY 4225 EDAGTVSFHLGNHASSAQLTVRAPEVTILEPDLQVOLSEGDASFOCLRSRSGQEAR---4282
DB 516 EDAGHYAVRSGGQSLAEVLQEKLEVYQSIADLAVGAKQAVFKEV---SDENVRGV 572
QY 4283 WALGGVFLQANEMNDITVEQ-GTLHLTLKHTLEDAQTVSF--HVGTCSSAQLKVTAK 4339
DB 573 WLKNGKELVPD--NRKIVSHIGRVHKLITIDVTPADEADYSFVPEGFACNLSAKLHFMEV 630
QY 4340 NTVVRGLENVLEALEGGALFEQCSQPE---VAAHTWLDDEPVRTSENAEVVFFENGRL 4396
DB 631 K-----IDFVRQPPKHILDCGSGPTDTIVVTGNKRLLDVPI-SGDPAFTVWOK---661
QY 4397 HLLKLLNLRPQDCSRVTFLLAGDMVTSAFLTVRGRWRLLEIPLKNAVAVRAGAAQARFTCTLS 4456
DB 682 -----TVTQ3KKASAGHPDAPEDAGA-----703
QY 4457 EAVPVGEASVINGAAVQPPDSDWTVTADGSHQALLRSNAQPHAGEVTFACDVAASAR 4516
DB 704 -----DEEVVFKKLLCETEGRVRVETTKDRSVFTVEGAKEDEGGYVTVTKNPFGEQ 757
QY 4517 LT-----VLGLPDPPEAEVVAHSSHTVTLSSWAFPMDSGGGLCGYRVEKEGATGQWR--4570
DB 758 VNLTVKVIDPDAPAPAKISNVGDCSTQWEPFAYDGGQVPLGYILERKKKSYRWML 817
QY 4571 ---LCHLVPQCEVCDGLAPGTGYRFRVAAGVGVGAGEPVHLPQTVRLAEPKVPVPOP 4627

```

RESULT 10

SN HUMAN

ID SN HUMAN

STANDARD;

PRT; 1709 AA.

AC Q9B222; Q96DL4; Q9GZS5; Q9H1H6; Q9H1H7; Q9H7L7;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Sialoadhesin precursor (Sialic acid binding Ig-like lectin-1) (Siglec-

DE 1) (CD169 antigen).

GN SN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.

RC TISSUE=monocytes;

RX MEDLINE=20575418; PubMed=11133773;

RA Hartnell A., Steel J., Turley H., Jones M., Jackson D.G.,

RA Crocker P.R.;

RT "Characterization of human sialoadhesin, a sialic acid binding

RT receptor expressed by resident and inflammatory macrophage

RT populations";

RL Blood 97:288-296(2001).

RN [2]

SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
MEDLINE=21639749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burdill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.V., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehaeslaio M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McClay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA Stuce C.D., Smith M.L., Soderlund C., Stewart C.A., Sulston J.E.,
RA Swann R.M., Sycamore A.C., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.,
RA "The DNA sequence and comparative analysis of human chromosome 20.";
RA Nature 414:865-871(2001).
[3]
SEQUENCE OF 733-1709 FROM N.A. (ISOFORMS 1 AND 2).
TIGS=Splice;
RC MEDLINE=21082933; PubMed=11214971;
RA Hattori A., Okumura K., Nagase T., Kikuno R., Hirosewa M., Ohara O.,
RA "Characterization of long cDNA clones from human adult spleen.";
RA DNA Res. 7:357-366(2000).
[4]
SEQUENCE OF 1539-1709 FROM N.A. (ISOFORM 3).
TISSUE=Thymus;
RC Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,
RA Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
RA Katsumura Y., Moriya S., Chiba E., Moniyama H., Onogawa S.,
RA Kariyama S., Satoh N., Matsunawa H., Takahashi E., Katsuka R.,
RA Kuga N., Kuroda A., Satoh I., Kanata K., Takami S., Terashima Y.,
RA Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Ota T.,
RA Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K.,
RA Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y.,
RA Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.,
RA "NEDO human cDNA sequencing project.";
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Macrophage-restricted adhesion molecule that mediates
CC sialic-acid dependent binding to lymphocytes, including
CC granulocytes, monocytes, natural killer cells, B-cells and CD8 T-
CC cells. Preferentially binds to alpha2,3-linked sialic acid (By
CC similarity). Binds to SPN/CD43 on T-cells (By similarity). May
CC play a role in hemopoiesis.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1) and
CC soluble (isoform 2).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9BZZ2-1; Sequence=VSP_002571;
CC Name=2;
CC IsoId=Q9BZZ2-2; Sequence=VSP_002571;
CC Name=3;
CC IsoId=Q9BZZ2-3; Sequence=VSP_002572;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Expressed by macrophages in various tissues.
CC High levels are found in spleen, lymph node, perivascular
CC macrophages in brain and lower levels in bone marrow, liver

CC Kupfer cells and lamina propria of colon and lung. Also expressed
CC by inflammatory macrophages in rheumatoid arthritis.
CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC
CC (sialic acid binding Ig-like lectin) family.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
CC -!- DATABASE: NAME=PROW; NOTE=PROW 2:18-22(2001);
CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/985165905.g.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF230073; AAK00757.1; -;
CC EMBL; AL109804; CAC17543.1; -;
CC EMBL; AL109804; CAC17542.1; -;
CC EMBL; AK024462; BAB15752.1; -;
CC EMBL; AK024459; BAB15749.1; -;
CC EMBL; AK024479; BAB15769.1; -;
CC EMBL; AK057560; BAB1527.1; -;
CC HSSP; Q62230; LQFO.
CC Genew; HGNC:11127; SN.
CC MIM; 600751; -;
CC GO; GO:0016021; C:integral to membrane; NAS.
CC GO; GO:000529; F:sugar binding; NAS.
CC GO; GO:0018337; P:cell-cell adhesion; NAS.
CC GO; GO:0007160; P:cell-matrix adhesion; NAS.
CC GO; GO:0006954; P:inflammatory response; NAS.
CC InterPro; IPR007110; IG-like.
CC InterPro; IPR003598; IG_c2.
CC Pfam; PF00047; ig; 14.
CC SMART; SM00408; IGC2; 4.
CC PROSITE; P550835; IG_LIKE; 14.
CC Cell adhesion; Lectin; Antigen; Transmembrane; Signal; Glycoprotein;
CC Immunoglobulin domain; Repeat; Alternative splicing; Polymorphism.
CC SIGNAL 1 19
CC BY SIMILARITY.
CC CHAIN 20 1709
CC EXTRACELLULAR (POTENTIAL).
CC DOMAIN 20 1641
CC TRANSMEM 1642 1662
CC DOMAIN 1663 1709
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC DOMAIN 20 136
CC IG-LIKE C2-TYPE 1.
CC DOMAIN 233 320
CC IG-LIKE C2-TYPE 2.
CC DOMAIN 326 405
CC IG-LIKE C2-TYPE 3.
CC DOMAIN 411 507
CC IG-LIKE C2-TYPE 4.
CC DOMAIN 511 593
CC IG-LIKE C2-TYPE 5.
CC DOMAIN 601 705
CC IG-LIKE C2-TYPE 6.
CC DOMAIN 708 785
CC IG-LIKE C2-TYPE 7.
CC DOMAIN 799 894
CC IG-LIKE C2-TYPE 8.
CC DOMAIN 898 977
CC IG-LIKE C2-TYPE 9.
CC DOMAIN 984 1083
CC IG-LIKE C2-TYPE 10.
CC DOMAIN 1085 1165
CC IG-LIKE C2-TYPE 11.
CC DOMAIN 1176 1248
CC IG-LIKE C2-TYPE 12.
CC DOMAIN 1259 1341
CC IG-LIKE C2-TYPE 13.
CC DOMAIN 1350 1442
CC IG-LIKE C2-TYPE 14.
CC DOMAIN 1445 1528
CC IG-LIKE C2-TYPE 15.
CC DOMAIN 1536 1631
CC IG-LIKE C2-TYPE 16.
CC BY SIMILARITY.
CC DISULFID 36 166
CC BY SIMILARITY.
CC DISULFID 41 98
CC BY SIMILARITY.
CC DISULFID 160 217
CC BY SIMILARITY.
CC DISULFID 262 305
CC BY SIMILARITY.
CC DISULFID 346 390
CC BY SIMILARITY.
CC DISULFID 433 491
CC BY SIMILARITY.
CC DISULFID 531 575
CC BY SIMILARITY.
CC DISULFID 624 689
CC BY SIMILARITY.
CC DISULFID 729 774
CC BY SIMILARITY.
CC DISULFID 817 876
CC BY SIMILARITY.
CC DISULFID 916 960
CC BY SIMILARITY.
CC DISULFID 1005 1067
CC BY SIMILARITY.

DISULFID 1107 1149 BY SIMILARITY.
 FT DISULFID 1193 1241 BY SIMILARITY.
 FT DISULFID 1281 1324 BY SIMILARITY.
 FT DISULFID 1367 1425 BY SIMILARITY.
 FT DISULFID 1465 1511 BY SIMILARITY.
 FT DISULFID 1554 1613 BY SIMILARITY.
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 265 265 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 339 339 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 697 697 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 726 726 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 730 730 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 741 741 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 886 886 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1138 1138 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1251 1251 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1462 1462 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1476 1476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 1632 1709 OSMGSEVMAFOKETTQIDPDPAATCETSTCAPLG ->
 FT FT GEGGLHPLGHSQKPS (in isoform 2).
 FT FT FTID=VSP_002571.
 FT FT RRRRVCKOSMGSEVMAFOKETTQIDPDPAATCETSTCAP
 FT FT PLG -> SSILMQPHVRPQVPHPWADQWCCFLPSGGESQ
 FT FT NL (in isoform 3).
 FT FT FTID=VSP_002572.
 FT FT K -> R (in dbSNP:625372).
 FT FT FTID=VAR_014136.
 FT FT H -> P (in dbSNP:709012).
 FT FT
 VARSPLIC 1666 1709
 VARIANT 239 239
 VARIANT 919 919
 Query Match 1.4%; Score 572.5; DB 1; Length 1709;
 Best Local Similarity 23.0%; Pred. No. 3.2e-12;
 Matches 399; Conservative 187; Mismatches 650; Indels 501; Gaps 92;
 QY 2703 LVHDSVPDAGLVTCVHSGSEETEARVRVHDLHVIGTK-RLKT-----NEVLGESCSP 2756
 DB 100 LLLKDLQPEDSGSNFNFETSEVNRWSDVKGTLVTVEEPVPTIASPVELLEGEVDFN 159
 QY 2757 C-----VLSHESADPAMWTVGGKTVGSSRRFQATPQGRKIVLVREAPS--DA 2804
 DB 160 CSTPYVCLQEQVRLQWQGPAPR-----SVTFNSQKFEPTGVGH--LETLHVMNSQDH 211
 QY 2805 GEVV---FSVRG--LTSKASLIVRERPAAI---IKPLEQOWVAPGDVEURCELSR---A 2853
 DB 212 GRILRCOLSANHRAQSEIHLQVAPKGVKILLSP-SGRNIIPLGELVTLTCQVNSSYPA 270
 QY 2854 GTPVHLKDKRAIKSKYDVVCEGTWAMLVIRGASLKDAGEYTCVEVASKST-----AS 2908
 DB 271 VSSIKWLKD-GVRLQTKGV-----LHLPQAAMSAGVYTQCAENGVSGLSVSPIS 320
 QY 2909 LHVEEKANCSTBELTNLQVEKG-----TAVFTCKTEHPAAT---VTWRKGLLELRASG 2959
 DB 321 LHI-----FMAEVQVSPAGPILFNQTVLVCNTPNEAPSDLRYSWYKNHVLLEDAH 371
 QY 2960 KHQPSQSGTLRLTISALEKADSDVTCDIGQAGSRAQLLVQGRRHITLEDLEDVDVQEG 3019
 DB 372 SH-----TLRLHLAT--RADTGFVCEV-----QNVHSGERSGPPSV--- 406
 QY 3020 SSATFRCRISPANYEPVHWFDKTP-----LHANELNE-----IDAQPGGY 3060
 DB 407 -----VWNHPL-----TPVLTAFLQAGLVGLHCVSVVEPLATLVLSHGG- 449
 QY 3061 HVLIT-----LROLAKDSGTIYFAGDORASA----- 3087
 DB 450 HILASTSGSDHSPRGSGTSGPNSLREIRDEETSDGEYKCSATNSLGNATSTLDFHAN 509
 QY 3088 ALRVTEKPSVFSRELTDAITEGEDTLVCETS---TCDIPMCWTWKDGTKLRGSARCQLS 3144
 DB 510 AARLLISPA-----AEVVEQAVTLCRSGLSPTPDARFSWYINGALL----- 552
 QY 3145 HECHRAQLLITGATLQDSGRYKC-----EAGGACSSSIVRV--HARVRFQEAALKDLEV 3196

DB 553 HEGPGSSLLPLAASSTDAGSVHCRARDGHSAGSPSPAVLTVLPVPPROPTTTL-DUDA 611
 QY 3197 LEGGA---ATLRC-VLSSVAAPVKWYGNVLRP-----GDKYSLRQEGAM 3238
 DB 612 ACAGAGRRGILLCRVSDPPPARLQLLHLDVRVATSLPSGGCGSTCGGCSPRMKVTKAENL 671
 QY 3239 LELVVRNLRPDQSGRYSC-----SFGDQTSATL-----TVTAL-PAOFTGKLRNKEATEGA 3289
 DB 672 LRVEITHNPLLEEGLYLCEASNALGNASTATFNGQATVLAIPSHTL-----QEGT 723
 QY 3290 TATLRCELSKTAPEVWRKSGSET---LRDGRYCLRQDQAMCELOIRGLAMVDAEYSCV 3345
 DB 724 EANTLCNVSREA-----AGSPANFSWFNG---VLMAQGPLETVTLLPVARTDAALYACR 775
 QY 3346 CGEE-----RTSASL-----TIRPMPAHFIFGLRHQES 3373
 DB 776 ILTEAGAQLSTFVLLSVLYPPDPKLSALLDMGQGHMAFICTVDSRPLALLA-LFHGEH 834
 QY 3374 IEGATATLRCELSKAAPVWRKGRESLRDGRHSLRODGAVCELOICGLAVADAGEYSC- 3432
 DB 835 L-----LATSGLQVPP-----SHGRFQAKAEANSLKLEVRGLGDSGVRCB 877
 QY 3433 ---VCGEERTSATLTVKALPAKFTGLRNEEAEVAGATAMLWCELSKVAP-----VEMRKG 3485
 DB 878 ATNVLGSSNTSLFQVR---GAWVQVSPSPQLQGGQAVLSCQVHTGVPEGTSYRW--- 930
 QY 3486 ENLRDGRYILRQSGTRCELOICGLAMADAGEYLCVCGOERTSATLIRA-----LPA 3538
 DB 931 ---YRDGQPL-----QESTSATLFAAITLTQAGYHCQ-AQAPGSATTSLAAPISLHVSYAP 984
 QY 3539 RFIEDVRKNQEAEGATAVLOCELNSAAPVWRKSGSETLRDGRY---SLRQDG----- 3588
 DB 985 RHVTLTLMDDTPGRGLGLLLCRVSDPPAQLR-----LLHGDRLVASTLQGVGPGESSP 1039
 QY 3589 -----TKCELOIRGLAMADTGEYSC---VCGOERTSAMLTVRALPKFTFEGLEGRNEE 3636
 DB 1040 RLHVAVAPNTURLIEHGMLEDEGVYICEASNTLQASASADFDAQAVNVQVWPG---AT 1096
 QY 3637 ATEGATAVLRCELSKMAPE---WVK-GHETLRDGRHSLRODGAUCELOIRGLVAEDA 3691
 DB 1097 VREGVNLVTLCLVWTHPAQLTYTWYDGGQRL---DAHS-----IPLPNVTRDA 1144
 QY 3692 GEYLCMOCKERTSAMLTVRAMPSKFIIEGLRN-----EATEGDTATLWCELSKAAPVW 3745
 DB 1145 TSYRCGVGPGRAPRLS-RPITLDVLYAPRNRLTYLLESHGGQALVLCVDSRPPAQL 1203
 QY 3746 RKGH-----ETLRDGRHSLRQDGSRCCELOIRGLAVVDAGYSCVC---G 3787
 DB 1204 ALSHAGRLLASSTAASVNTLR-----LELRGPQPRDEGYSCARSPLG 1248
 QY 3788 QERTSATLTVRALPARFIEDVKMQEAR--EGATAVLOC-ELSKAAPV--EWRKSGSETLRG 3842
 DB 1249 QANTSLERLEG-----VRVILAPEAAVPEGAPITVTTCADPAHAAPTLYTWYHNGRWL-- 1301
 QY 3843 GDRYSLRQDGTGRCELOIHGLSVADTGEYSCVC-----GQERTSATLTVPAQPVFREP 3896
 DB 1302 -----QEGPAASLSFLVATRAHAGAYSQAQADQGRSSRPAALQVLYAPQDAVLUSF 1354
 QY 3897 QSLQAEEGSTATLOCEL-SEPTA-----TVWMSKGLQLOANGERRERPRLOQTAELV 3947
 DB 1355 RDSRAR--SNAVICTVDSEPPAEALSHDGKVLATSSGVHSLASGTGHVQVARNALRLQ 1412
 QY 3948 LODLQREDTGEYTCCTC-----GSCATSATLTVTAAPVRFLELOHQVDEGGTAHLCELS 4003
 DB 1413 VQDVPAGD-DTYVCTAQNLLGSIITGLRQVEGA--RVVAE-PGLDVPEGAALNLSRLL 1468
 QY 4004 RAGASVE-----WRKGSLLQFPCKAKYQMVQGAALLLVRGVEQEDADGYTC---DT 4052
 DB 1469 GGPVGVGNSTFAWFWNDRRLHAEF-----VPTLAFTHVARAQAGWYHCLAEPLT 1517
 QY 4053 GHTQSM-ASLSVRVPRPKFKTRLOSLEOETGDIARLCCQLSDAESGAIVQWLKEGVELHA 4111

Db 1518 GAASAPVMLRLVLP-PKTPMMVFVEGGLRGILDCRV-DSEPLA-----SUTLHL 1568
 QY 4112 GPKYEMRS--QGATRELLIH-----QLEAKDTGEYAC-----VTGGOKTA 4149
 Db 1569 GSRUVASSQPGCAPAEPHIVLSPNALRVVDIEALRPSDQGEVYCSASNVLSASTS 1625

RESULT 11

MYPC HUMAN STANDARD; PRT; 1274 AA.
 AC Q14856; Q16410; Q9UE27;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Myosin-binding protein C, cardiac-type (Cardiac MyBP-C) (C-protein,
 DE cardiac muscle isoform).
 GN MYBPC3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=95262635; PubMed=7744002;
 RA Gautel M., Zuffardi O., Freiburg A., Labeit S.;
 RT "Phosphorylation switches specific for the cardiac isoform of myosin
 RT binding protein-C: a modulator of cardiac contraction?";
 RL EMO J. 14:1952-1960(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANT CMH4 GLN-542.
 RX MEDLINE=97200835; PubMed=9048664;
 RA Carrier L., Bonne G., Bahrend E., Yu B., Richard P., Niel F.,
 RA Hainque B., Craud C., Gary F., Labeit S., Bouchour J.-S., Dubourg O.,
 RA Desnos M., Hagege A.A., Trent R.J., Komajda M., Fiszman M.,
 RA Schwartz K.;
 RT "Organization and sequence of human cardiac myosin binding protein c
 RT gene (MYBPC3) and identification of mutations predicted to produce
 RT truncated proteins in familial hypertrophic cardiomyopathy.";
 RL Circ. Res. 80:427-434(1997).
 RN [3]
 RP SEQUENCE OF 640-694 FROM N.A.
 RX MEDLINE=96083593; PubMed=7493026;
 RA Bonne G., Carrier L., Bercovich J., Craud C., Richard P., Hainque B.,
 RA Gautel M., Labeit S., James M., Beckmann J.;
 RT "Cardiac myosin binding protein-C gene splice acceptor site mutation
 RT is associated with familial hypertrophic cardiomyopathy.";
 RL Nat. Genet. 11:438-440(1995).
 RN [4]
 RP VARIANT CMH4 LYS-755.
 RX MEDLINE=98200174; PubMed=9541104;
 RA Yu B., French J.A., Carrier L., Jeremy R.W., McTaggart D.R.,
 RA Nicholson M.R., Hamblly B., Sensarian C., Richmond D.R., Schwartz K.,
 RA Trent R.J.;
 RT "Molecular pathology of familial hypertrophic cardiomyopathy caused by
 RT mutations in the cardiac myosin binding protein C gene.";
 RL J. Med. Genet. 35:205-210(1998).
 RN [5]
 RP VARIANT CMH4 HIS-654.
 RX MEDLINE=98200185; PubMed=9541115;
 RA Woolman-Smoak J.C., Mayosi B., Brink P., Corfield V.A.;
 RT "Identification of a new missense mutation in MYBPC-C associated with
 RT hypertrophic cardiomyopathy.";
 RL J. Med. Genet. 35:253-254(1998).
 CC -1- FUNCTION: THICK FILAMENT-ASSOCIATED PROTEIN LOCATED IN THE
 CC CROSSBRIDGE REGION OF VERTEBRATE STRIATED MUSCLE. IN VITRO
 CC IT BINDS MHC, F-ACTIN AND NATIVE THIN FILAMENTS, AND MODIFIES THE
 CC ACTIVITY OF ACTIN-ACTIVED MYOSIN ATPASE. IT MAY MODULATE MUSCLE
 CC CONTRACTION OR MAY PLAY A MORE STRUCTURAL ROLE.
 CC -1- PTM: SUBSTRATE FOR PHOSPHORYLATION BY PKA AND PKC. REVERSIBLE
 CC PHOSPHORYLATION APPEARS TO MODULATE CONTRACTION (BY SIMILARITY).
 CC -1- DISEASE: Defects in MYBPC3 are the cause of familial hypertrophic
 CC cardiomyopathy type 4 (CMH4) [MIM:115197]; also known as FHC type

CC 4. CMH4 is an autosomal dominant disorder. It is a primary cardiac
 CC disease with risk of sudden unexpected death and is characterized
 CC by an unexplained hypertrophied non-dilated ventricle and by
 CC myocytic and myofibrillar disarray.
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. MyBP
 CC family.
 CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 3 fibronectin type III domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X84075; CAA58882.1; -;
 CC EMBL; Y10129; CAA71216.1; -;
 CC EMBL; S80778; AAB35662.1; -;
 CC FIR; S55050; S55050.
 CC DR HSSP; P56276; ITLK.
 CC DR Genew; HGNC:7551; MYBPC3.
 CC DR MIM; 600958; -;
 CC DR MIM; 115197; -;
 CC GO; GO:0008307; F-structural constituent of muscle; TAS.
 CC GO; GO:0008936; P-muscle contraction; TAS.
 CC GO; GO:0008942; P-regulation of striated muscle contraction; TAS.
 CC InterPro; IPR008957; FN III-like.
 CC InterPro; IPR003961; FN III.
 CC InterPro; IPR007110; IG-like.
 CC InterPro; IPR003598; IG_c2.
 CC Pfam; PF00041; f3; 3.
 CC Pfam; PF00047; ig1; 7.
 CC SMART; SM00060; FN3; 3.
 CC SMART; SM00408; IGC2; 1.
 CC PROSITE; PS50835; IG LIKE; 6.
 CC Immunoglobulin domain; Actin-binding; Cell adhesion; Muscle protein;
 CC Thick filament; Repeat; Phosphorylation; Disease mutation;
 CC Cardiomyopathy; Polymorphism.
 CC FT DOMAIN 102 152
 CC FT DOMAIN 153 256
 CC FT DOMAIN 257 352
 CC FT DOMAIN 353 452
 CC FT DOMAIN 453 543
 CC FT DOMAIN 544 633
 CC FT DOMAIN 634 771
 CC FT DOMAIN 772 869
 CC FT DOMAIN 870 970
 CC FT DOMAIN 971 1085
 CC FT DOMAIN 1086 1180
 CC FT DOMAIN 1181 1274
 CC MOD_RES 275 275
 CC MOD_RES 284 284
 CC MOD_RES 304 304
 CC MOD_RES 304 304
 CC VARIANT 542 542
 CC VARIANT 654 654
 CC VARIANT 755 755
 CC VARIANT 536 536
 CC CONFLICT 536 536
 CC SEQUENCE 1274 AA; 140833 MW; D31FOEA1572E5417 CRC64;
 CC
 CC Query Match 1.4%; Score 571; DB 1; Length 1274;
 CC Best Local Similarity 21.3%; Pred. No. 2.5e-12;
 CC Matches 336; Conservative 176; Mismatches 585; Indels 478; Gaps 63;
 CC
 CC QY 3804 FIEDVKNQREAGATVLCQELSKA-APVEMKSGSETLGGDRYSLRQDGTCELIHGL 3862
 CC Db 12 FSKPKRSVEAAGSPAVFEATERAGVKVWRQGRGSDISASNKYGLGATEGTRHLLTVREV 71

RA Kelm S., Pelz A., Schauer R., Filbin M.T., Tang S., de Bellard M.E.,
RA Schnaar R.L., Mahoney J.A., Hartnell A., Bradfield P., Crocker P.R.;
RT "Sialoadhesin, myelin-associated glycoprotein and CD22 define a new
RT family of sialic acid-dependent adhesion molecules of the
RL immunoglobulin superfamily";
RL Curr. Biol. 4:965-972(1994).
RN BINDING TO SPN.
RP MEDLINE=21136329; PubMed=12338599;
RA van den Berg T.K., Nath D., Ziltener H.J., Vestweber D., Fukuda M.,
RA "CD43 functions as a T cell counterreceptor for the macrophage
RT adhesion receptor sialoadhesin (siglec-1).";
RL J. Immunol. 166:3637-3640(2001).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 20-138.
RA MEDLINE=98325385; PubMed=9660955;
RA May A.P., Robinson R.C., Vinson M., Crocker P.R., Jones E.Y.;
RT "Crystal structure of the N-terminal domain of sialoadhesin in
RT complex with 3' sialyllactose at 1.85 A resolution.";
RL Mol. Cell 1:719-728(1998).
RN [7]
RP STRUCTURE OF 20-138 BY NMR, AND MUTAGENESIS OF TRP-21 AND ARG-116.
RA MEDLINE=99321481; PubMed=10393093;
RA Crocker P.R., Vinson M., Kelm S., Drickamer K.;
RT "Molecular analysis of sialoside binding to sialoadhesin by NMR and
RT site-directed mutagenesis";
RL Biochem. J. 341:355-361(1999).
CC -!- FUNCTION: Macrophage-restricted adhesion molecule that mediates
CC sialic-acid dependent binding to lymphocytes, including
CC granulocytes, monocytes, natural killer cells, B-cells and CD8 T-
CC cells (By similarity). Preferentially binds to alpha2,3-linked
CC sialic acid. Binds to SPN/CD43 on T-cells. May play a role in
CC hemopoiesis.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1) and
CC soluble (isoforms 2 and 3).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q62230-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q62230-2; Sequence=VSP_002573, VSP_002574;
CC Name=3;
CC IsoId=Q62230-3; Sequence=VSP_002575, VSP_002576;
CC -!- TISSUE SPECIFICITY: Expressed by macrophages in various tissues.
CC Highest expression in spleen and lymph node with lower amounts in
CC lung, liver, bone marrow, heart and skin. No expression in thymus,
CC kidney, brain or small intestine.
CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC
CC (sialic acid binding Ig-like lectin) family.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z36293; CA855290.1; -;
DR EMBL; Z36233; CA855268.1; -;
DR EMBL; Z36234; CA855269.1; -;
DR EMBL; U92842; AAB95641.1; -;
DR EMBL; U92833; AAB95641.1; JOINED.
DR EMBL; U92834; AAB95641.1; JOINED.
DR EMBL; U92836; AAB95641.1; JOINED.
DR EMBL; U92837; AAB95641.1; JOINED.
DR EMBL; U92838; AAB95641.1; JOINED.
DR EMBL; U92839; AAB95641.1; JOINED.
DR EMBL; U92840; AAB95641.1; JOINED.
DR EMBL; U92841; AAB95641.1; JOINED.

DR PIR; S50065; S50065.
DR PDB; 1QFO; 16-APR-99.
DR PDB; 1QFP; 16-APR-99.
DR MGD; MGI:99668; Sn.
DR GO; GO:0016021; C:integral to membrane; ISS.
DR GO; GO:0005529; F:sugar binding; ISS.
DR GO; GO:0018337; P:cell-cell adhesion; ISS.
DR GO; GO:0007160; P:cell-matrix adhesion; ISS.
DR GO; GO:0006954; P:inflammatory response; ISS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00047; ig; 16.
DR SMART; SM00408; IGc2; 5.
DR PROSITE; PS00835; IG-LIKE; 14.
DR PROSITE; PS00290; IG_MHC; 1.
KW Cell adhesion; Lectin; Transmembrane; Signal; Glycoprotein;
KW Immunoglobulin domain; Repeat; Alternative splicing; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 1694 SIALOADHESIN (POTENTIAL).
FT DOMAIN 20 1638 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1639 1659 POTENTIAL.
FT DOMAIN 1660 1694 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 136 IG-LIKE V-TYPE.
FT DOMAIN 153 235 IG-LIKE C2-TYPE 1.
FT DOMAIN 239 321 IG-LIKE C2-TYPE 2.
FT DOMAIN 326 406 IG-LIKE C2-TYPE 3.
FT DOMAIN 416 508 IG-LIKE C2-TYPE 4.
FT DOMAIN 509 594 IG-LIKE C2-TYPE 5.
FT DOMAIN 602 701 IG-LIKE C2-TYPE 6.
FT DOMAIN 704 781 IG-LIKE C2-TYPE 7.
FT DOMAIN 795 890 IG-LIKE C2-TYPE 8.
FT DOMAIN 894 972 IG-LIKE C2-TYPE 9.
FT DOMAIN 979 1078 IG-LIKE C2-TYPE 10.
FT DOMAIN 1080 1160 IG-LIKE C2-TYPE 11.
FT DOMAIN 1171 1263 IG-LIKE C2-TYPE 12.
FT DOMAIN 1244 1336 IG-LIKE C2-TYPE 13.
FT DOMAIN 1341 1438 IG-LIKE C2-TYPE 14.
FT DOMAIN 1441 1519 IG-LIKE C2-TYPE 15.
FT DOMAIN 1533 1626 IG-LIKE C2-TYPE 16.
FT SITE 827 829 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 36 166 BY SIMILARITY.
FT DISULFID 41 98 BY SIMILARITY.
FT DISULFID 160 218 BY SIMILARITY.
FT DISULFID 263 306 BY SIMILARITY.
FT DISULFID 347 391 BY SIMILARITY.
FT DISULFID 434 492 BY SIMILARITY.
FT DISULFID 532 576 BY SIMILARITY.
FT DISULFID 625 685 BY SIMILARITY.
FT DISULFID 725 770 BY SIMILARITY.
FT DISULFID 813 872 BY SIMILARITY.
FT DISULFID 911 955 BY SIMILARITY.
FT DISULFID 1000 1062 BY SIMILARITY.
FT DISULFID 1102 1144 BY SIMILARITY.
FT DISULFID 1188 1236 BY SIMILARITY.
FT DISULFID 1276 1319 BY SIMILARITY.
FT DISULFID 1362 1421 BY SIMILARITY.
FT DISULFID 1462 1508 BY SIMILARITY.
FT DISULFID 1551 1610 BY SIMILARITY.
FT CARBOHYD 159 159 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 583 583 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1459 1459 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1473 1473 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPLIC 326 340 MAEVKMPAGVLEN -> SESWMLRGVPVSGKH (in
isoform 2). /FtId=VSP_002573.
FT VARSPLIC 341 1694 Missing (in isoform 2). /FtId=VSP_002574.
FT VARSPLIC 1528 1598 LVASNLHDAKPHIRVAPPNALRVIE -> CEVEPIS
ALCLSLHGTGPYQAFSSAQSGKGLRFLASSLAGCMWF
VSMGLYPALKWRILLPWFDEYRR (in isoform 3).
FT VARSPLIC 1599 1694 Missing (in isoform 3). /FtId=VSP_002575.
FT MUTAGEN 21 21 W->O: LOSS OF SIALIC ACID BINDING.
FT MUTAGEN 116 116 R->A: LOSS OF SIALIC ACID BINDING.
FT MUTAGEN 116 116 R->L: 10-FOLD LOSS IN AFFINITY TO SIALIC

Query Match 1.3%; Score 532; DB 1; Length 1694;
Best Local Similarity 20.7%; Pred. No. 7.5e-11;
Matches 416; Conservative 240; Mismatches 691; Indels 660; Gaps 100;

QY 790 SVQGLARFLKDMAGSC--VDVAGGPAQFECETSEAHVHVHYKD---GMLGHSGER 843
DB 28 NVQGL-----SGSCLLIPCISYPADVPVNG---ITAIWYDYSGRQVVIHSGDP 76
QY 844 FLOEDVGRHR-----LVAATVTRQDEGTYSYCRVGBDSVDFLRVSE--- 885
DB 77 KL--VDKRFGRAGELMGNHDKVNCNLLDLKPEDSGTY-----NFRFEISDSNR 124
QY 886 -----PKVYFAKQLARRKLOAEAGASATLSCEVAQAQTEVTWY--KDGK 928
DB 125 WLDVKGTTVTTDPSPPTTIPBEL-REGNERNFNCSTPYLC-LOEKQVSLQWRQDPT 182
QY 929 KLSSSSKYCMBATCTRLRVVQQA-----GQADAGEYSCAGGQRLSFHLDYKE-PKVY 981
DB 183 HSVTSSQSLPTSGYHQTTLHMSLWQDHGRTLLCQFSLGAHSSRKEVVLQVPHAPKV 242
QY 982 FAKQOVAHSEVOAEGANATLSCEVAO---AQAEVWYKDGKLSLKVHVEAKGRRR 1038
DB 243 --EILLSSGNILPGDPVTLTKWNSSYPAVSAVOWARDGVNL--GVTHGV----- 290
QY 1039 LVVQOAGKTADGYSCEA---RGQVRS--FRLHTEPKMFAKEQS VHVNEVOAEGASAM 1093
DB 291 LRLFAANWDSGATCOATNDMSGLVSSPLSLHV-----FMAEVKMNPAQVLENETVT 344
QY 1094 LSCEV---AQAEVWYKDGKLSLSSKVMGVKGTRELVLPOAKADAGEYSCBA-- 1148
DB 345 LLCSTPKAEPOELRYSYKHNHILLEDAAH-----ASTLHLPVAVTRADTGFYFCEVQN 395
QY 1149 --GGQVRSFHLHTEPKGVFAKEQS VHVNE-----VOAEAGTAMLSCE-VAQPQTEVTWYK 1201
DB 396 AQGERS-----PLSVVRYPLPTDPTTLETOAGLVGLHCSVSEPLATVLSH 448
QY 1202 DGKLSSSS-----KYRMEVKGCTRLRVVQVKGADAGEYSCBA-----GGQVRSFQLH 1250
DB 449 GGLTLASNGENDFNPRF--SSAPNSLRLEIRDLQPADSGEYCTCLAVNSLGNSTSLDFY 508
QY 1251 ITEPKAVFAKEQVHVNEVTEAGASATLSCEVAQA---QTEVTWYKDGKL---SSSSKV 1304
DB 509 ANVARL-----LNPAAVEVQGAVTLSCRSGLSPADPTFRFSWYINGALLLEGSSSS-- 560
QY 1305 RIEAAGCMRLVVOQAQADAGEYTC--EAG-----LALLPAASSTDAGSYCYCTQAGPNTSGPSLPTVTFYPPRPKPTFTARLDL 1333
DB 561 -----LRLFAANWDSGATCOATNDMSGLVSSPLSLHV-----FMAEVKMNPAQVLENETVT 344
QY 1334 -----GQR--LSFHLVDSEPKAVFAKEQLAHKKVQAEAGATATLS-----CEVAQAQTE 1380
DB 611 DTSGVGGRGILLCHVDSDPP-----AQLRLHKK-----GHVATSLPSRCSCSQRTK 660
QY 1381 VTWYKDGKLSLSSSKVAMEAVGCTRLRVVQQAQADAGEYSCBA-----GGQVRSFSLDVA 1436
DB 661 VS-----RTSNLSHVE-----IQKVLDESGVTLCEASNTLGNSSAAASFNAK 703

QY 1437 BPKVVFKEQPVHREVQQAAGASTTLC-----EVAQAQTEVMYKDGKLSLSSSKVME 1492
DB 704 ATVLVITSNTL-----REGTEANLTCNGQEVAVSPANFSWFRNGV-----LMT 748
QY 1493 VGCTRLRVVQQAQADAGEYSC-----EAGSQRLSFHLHVAEP---KAVFAKEQ-ASREV 1544
DB 749 QGSLETVRLLQARTDAVAAACRLLTEDGAQ-----LSAPVVLVLYADPPKLSALL 801
QY 1545 QAEAGTSTLSCEV-AQAQTEVTWYKDGKLS-----SSSKVMEAVGCTRLRVVQ 1595
DB 802 DVGQGHMAVFICTVDVSYPLAHLISFRGHLLATNLLEPORPSHGRIQAKATANSIQLEVRE 861
QY 1596 AQQAQADAGEYSCA-----GQRLSFHLHVAEPKVFAKEQPAHREVQAEAGASATLSCEVA 1651
DB 862 LGLVDGNGVHCHEATNILGSANSSLFQVGAWVFT-----ITELREGQAVVLSQV 914
QY 1652 QAEQTEVT---WYKDGKLS-----SSSKVMEAVGCTRLRVVQQAQADAGEYSCBAGGQRLS 1706
DB 915 TGVSEGTYSYQDGRPLCESTSTLRIAAISLRQ-----AGAVHCQA----- 957
QY 1707 FRLHVAELPEQISERPCRRREPLVVKHEHDIILTATLATPSAATVTLKDGVEIRSRKHE 1766
DB 958 -----QAPDTAI-ASLAAPVSLHVS-----TERHV 982
QY 1767 TASQGDTHLTVHGAQVLDLSAIYSCRVAEGQDPFQVVEVAAKFCLLPEVCGELGTV 1826
DB 983 TLS-----ALLSTDPERLGHVCSVQ-----SDPPAQLQFHRN--RL---VASTLQAD 1027
QY 1827 TLACELSPACAEVWRCNGTQPRVKRFQVMAEGVRSLSVL--GLR-----AEDAGE 1877
DB 1028 ELA-----GSNPL-----HVTLPNELQLQHFPLEDDGT 1059
QY 1878 YVCESRDHTSAQLTVSVRVVVKFMSGLSTVVASGEBAFQCWV---SPSDVAVWFRD 1934
DB 1060 YTCESANTLGASAAADPAQAVRVVMPNATVQEGQVNLTCVWSTHODLSYTYK 1119
QY 1935 GALLQSEKFAISQSGASHSLTISDLVLEDAG--QITVEAEGASSAALRVREAPVLFK 1992
DB 1120 G-----QOLLGARSITLPSVKVLDATSVRCGVGLPG-----HAPHLSR- 1157
QY 1993 KLEPQTVBEERSVTLLEVELTFPPELWTRN--ATALAPGNVBIHAEGARHLRVLHVNG 2050
DB 1158 -----PVTLDV-----LHAPNRLTYL-----LETQGRQLALVL--- 1187
QY 2051 FADRGFFCEFPDXTQAKLTVMQVRLVYRGLQAVEAREQGT--TMEVQLSHADVGS 2108
DB 1188 -----C-TVDSRPPAQLT-----LSHGDQDVASSTASVPTNRLUEL----- 1223
QY 2109 WTRDGLRFQOQPTCHLAVRGPMTLTLGLRPEDSGMLVFKAEG-----VHTSARLVVTEL 2164
DB 1224 -----QDP-----RPSNEGLYSCSAHSPGLKANTISLELLEGV 1256
QY 2165 PVSFRLQDVVTEKEKVTLECE--LSRPNVVRWLKDGVELPAGTKMAIAQACRSL 2222
DB 1257 RYQMN-PGSGV--PEGEPTVTCEDPAALSSALYAFHNGHWLQ-----EGPASSL 1304
QY 2223 TIYRCEFAQGVVYCDADHAQSSASVKVQGRTYTLIYRRLVLAEDAGEIQFVAENAEARQ 2282
DB 1305 QFLVTRAHAGAYFCQVHDTQGTSSR--PASLQILY-----APRDAVLSSERDSR 1355
QY 2283 LRV-----KELPVTILVRPLRDKIAMEKHR-----GVLEQVSRASAOVWFKGSOEL 2329
DB 1356 LMWVTOCTVDSEPPAEVMSLHNGKVAASHERHSASGIGHIQVARNAL----- 1404
QY 2330 QGPKYELVSDGLYKLIISDVHAEDEDTYTCDA---GDVKTSAQFFVEEQSITIVRGL 2385
DB 1405 -----RLQVQVTLGDGNTYVCTAQHTLGSISTTQRLITE-----TDIR-- 1443
QY 2386 QDVTWEPAPAWFCESTISPSVRPKMLLGKTVLQAGN----- 2424
DB 1444 --VTAEPGLDW-----PEGTALNLSCLLPGSGPTGNSFTFWNRHRLHAP 1489
QY 2425 -----VGLEQSGTVHRLMLRITCTMTGTPVHFTVGKSSARLVVSDIPVLTPL 2476

```

Db 1490 VPTLSFVPTVRAQAGLHCRADLTGATTSAPVWMLRVLYPDKPTLIVF-----VE 1540
Qy 2477 PXTGRELSVSVLSCDFRPAKAVQWYK-----DTPLSPSKPKXNSLEGQWAEAL 2525
Db 1541 PQGGHQ---GILDCRVDSPLAILTHRGSLVASNLQHDAPTKP--HIRVTAPPNALRV 1595
Qy 2526 RILMLPADAGVRCQA---GSAHSS 2548
Db 1596 DIEELGFSNQGEYVCTASNLUGSASAS 1622

RESULT 23
MYBP CHICK
ID MYBP CHICK STANDARD; PRT; 1131 AA.
AC P16419, Q90606;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin-binding protein C, fast-type (Fast MyBP-C) (C-protein, skeletal
  muscle fast-isoform).
GN MYBP2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-24; 182-198 AND 1030-1044.
RC TISSUE=Skeletal muscle;
RX MEDLINE=9404344; PubMed=8227129;
RA Okagaki T., Weber F.E., Fischman D.A., Vaughan K.T., Mikawa T.,
  Reinach F.C.;
RT "The major myosin-binding domain of skeletal muscle MyBP-C (C
  protein) resides in the COOH-terminal, immunoglobulin C2 motif.";
RL J. Cell Biol. 123:619-626(1993).
RN [2]
RP SEQUENCE OF:139-1131 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=90192766; PubMed=2315308;
RA Einheber S., Fischman D.A.;
RT "Isolation and characterization of a cDNA clone encoding avian
  skeletal muscle C-protein; an intracellular member of the
  immunoglobulin superfamily.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2157-2161(1990).
CC -!- FUNCTION: THICK FILAMENT-ASSOCIATED PROTEIN LOCATED IN THE
  CROSSBRIDGE REGION OF VERTEBRATE STRIATED MUSCLE A BANDS. IN VITRO
  IT BINDS MHC, F-ACTIN AND NATIVE THIN FILAMENTS, AND MODIFIES THE
  ACTIVITY OF ACTIN-ACTIVED MYOSIN ATPASE. IT MAY MODULATE MUSCLE
  CONTRACTION OR MAY PLAY A MORE STRUCTURAL ROLE.
CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. MyBP
  family.
CC -!- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00922; AAC59644.1; -
CC EMBL; U00923; AAC04307.1; -
CC EMBL; M31209; AAA49068.1; -
CC PIR; A48668; A35089.
CC RSP; P56276; 1TLK.
CC InterPro; IPR008957; FN III-like.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR003962; FN_III_subd.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003598; Ig_c2.

```

```

DR Pfam; PF00041; fn3; 3.
DR Pfam; PF00047; ig; 7.
DR PRINTS; PRO0014; FNTYPEIII.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00408; IGc2; 1.
DR PROSITE; PS00835; IG LIKE; 5.
KW Immunoglobulin domain; Actin-binding; Cell adhesion; Muscle protein;
KW Thick filament; Repeat.
FT INIT MET 0
FT DOMAIN 47 148 IG-LIKE C2-TYPE 1.
FT DOMAIN 248 337 IG-LIKE C2-TYPE 2.
FT DOMAIN 338 428 IG-LIKE C2-TYPE 3.
FT DOMAIN 429 529 IG-LIKE C2-TYPE 4.
FT DOMAIN 530 629 IG-LIKE C2-TYPE 5.
FT DOMAIN 630 727 FIBRONECTIN TYPE-III 1.
FT DOMAIN 728 828 FIBRONECTIN TYPE-III 2.
FT DOMAIN 829 922 IG-LIKE C2-TYPE 6.
FT DOMAIN 923 1037 FIBRONECTIN TYPE-III 3.
FT DOMAIN 1038 1131 IG-LIKE C2-TYPE 7.
SQ SEQUENCE 1131 AA; 126811 MW; 73262DC5A939D6AC CRC64;

Query Match 1.3%; Score 528.5; DB 1; Length 1131;
Best Local Similarity 21.0%; Pred. No. 66-11;
Matches 306; Conservative 172; Mismatches 523; Indels 453; Gaps 53;

Qy 3817 ATAVIQLCELSKAAPVWYK-----GSETLRGGDRYSLRQDGTTC-ELQIHGLSVADTGE 3869
Db 73 AGAALPC-----APAVKWFKGMWELGDKSARCLRHSDVDDDKVHTFELTITTKVAMGDRGD 128
Qy 3870 YSC-YGQER---TSATLTVRAPQVFPFQSL-QAEEGSTATLQCELSSEPTATVWVK 3924
Db 129 YCEVTAKEQKSCSFSDIAPRSEGNVLOAFKRTGEGKDDT-----AGELDFFGLLKK 184
Qy 3925 GGLQLQANGRRPR-----LQCTABLVLQDLQREDTGYTCTCGSQATSATLT 3973
Db 185 REVQVEEKKKKDEDDQPPPELWELKGVTKKSEVERI---AFQYGIIDLRGMLKRLKK 240
Qy 3974 VTAAPVR---FLRELQ-HQEVDEGGTAHLCCELSRAGASVEWEKSLQIFPCAKYQWQD 4029
Db 241 VHVPEKSEAFIRKLDPAQVQDKGNKILWELSDPLPLKWKYKNGQLLKPSTKYFENV 300
Qy 4030 GAAAEELVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRQLQSLQEQETGDIARLCC 4089
Db 301 GLKRLITHKSLADDAAYECVNDKCFTEVFVKEPPVTVVRGLEDQVVGVD--RVVL 358
Qy 4090 QLSDAESGAVQWMLKEGVELHAGP--KYEMRSQGTARELLIHQLEAKDTGEVACTVGGQK 4147
Db 359 EAEVSEGAQVWMLKDGVDVTRDDAFKRYFKKDKGKHFLLINEARLSDSAHYKIMTNGGE 418
Qy 4148 TAASLRVTEPVTIVRGLVDAEVTADEVERFSECVSRAGATGVQWCLQGLPLQSNVETEV 4207
Db 419 SEAELSVEEKQLEVLQDMADLTVKASQAVFKEVSDSEKVTG-RWFRNGVVEKPKRIHI 477
Qy 4208 AVRDGRIRHLRLKGVTPEDAGTVSPFLGNHASSAQLTVRAPEVTILEPQDVQLSEGQDA 4267
Db 478 S-HNGRFHKLVIDDVRPEDEGDTTIPGYALSLSAKLNFLKIV-----521
Qy 4268 SFQCLSRASGQEAQWALGVPLQANEMNDIIVEQGTLEHLLTLHKVLTLEDAGTVSFHVT 4327
Db 522 -----EYVPKQ-----EPPKHL-----D 535
Qy 4328 CSSEALQKVTAKNTVVRGLENVEALEGGEALFECOLSQPEVAHAHTWLLDDEPVRTSENAE 4387
Db 536 CSGKA-----AENTIV-----VVAGNKVRLDVISGEPAFTVTKSGDQLFTATE--- 580
Qy 4388 VVFFENGLRHLLLLKNLRPDQSCRVTFLAGDMVTSAFLTVRGWRLEILLEPLKNAAVRAGA 4447
Db 581 -----GRVHI-----DS-----QADL--SSFVIESAERSD-----603
Qy 4448 QARETCTLSEAVPVGEASWYINGAAVQPDSDTWTADGSHQALLLRSAQPHHAGEVTIFA 4507
Db 604 EGRYCITVTN--PVGEDS-----619

```


QY 4011 WRK-----GSLQLPCCAKYQVQGAALVLRVGEQDAGDYTCDTGHTQSM 4059
Db 226 YEKIAFOYGTDLGMLKRLKAKVKKSGAA----- 257
QY 4060 SLSVRPRPFKTRLSLEQ-ETGDIALRLCCQLSDAESGAVWQWLGKGVHAGPKYEMR 4118
Db 258 -----FTKLDPAQVDRGNKIKLWVEISDDP--LTLKFKNGQEIKESSKYVPE 305
QY 4119 SQGATRELLIHQLEAKTGBYACVTGGKTAASLRVTEPEVTVIRGLVDAEVTADDEVEF 4178
Db 306 NVGKRILITINKCTLADDAAYEAVKDEKCTELFVKPEPVLIVTPELDDQOVFGDREVM 365
QY 4179 SCEVSRAGATGVQVCLQGLPLQSNVEVAVR--DGIHTLRLKGVTPEDAGTVSFHLG 4235
Db 366 AVEVSEGAQ-VMMKDGVELTREDSPKARYRFKDGKRLHILFSDVVQEDRGYQVITN 424
QY 4236 NHASSAQLTVRAPEVTILEPLQVQLSEGQDASQCRLS--RASGQEARWALGGVPLQAN 4293
Db 425 GGOCEABELIVEKQLEVLQDIADLTVKASEQAVFKCVSDKVTG---KWKNGVEVRPS 481
QY 4294 EMNDITVEQ-GTLHLTLHLKVTLEDAQTVSF-----HVGTCSE--AQLKVTAKTVVR 4344
Db 482 KR--ITISHVRPHKLVIDVRPEDEGDTYFVPGYALGSLKAKINLEIKV----- 531
QY 4345 GLENVEALEGGEALFEQQLSQPEVAATHLLDDPEVPTSENAEVVFPENGRLHLLKLN 4404
Db 532 --EYVQKPEPKIHLDCS-----GXTSENAIVVAGNKLRL--LDVSI 569
QY 4405 RPQSCRVTFLAGDMVTSFVTVGWRLEILEPLKNAAVRAGAQAQRTCTLSEAVPGEA 4464
Db 570 TGEPPVATWLGDEV---FTTTEG-----RTRIEKRYDC----- 601
QY 4465 SWYINGAAVQDDSDTWTADGSHQALLRSQAQPHAGEVTFACRDVAVASARLTVLGLPD 4524
Db 602 SSFVIESAQREDEGRYII-----KVTNPGVEDVAFIFQVVDVDP 641
QY 4525 PPDAEVVAHSSHTVTLSSAAPMSDGGGLGCGYRVKVGATQW-RLCHELVGPPECVV 4583
Db 642 PPEAVRITSVGEDWAILWPEPMYDGGKFTVGYLVERKKGSQWMLNFEVTEITYES 701
QY 4584 DGLAPGETYRRAVAVPGVAGEPVHLPTQVRLAEPKPPVPPQSPAPESQVAGEDVSL 4643
Db 702 TKMIEGLIYERFAVNAIGVSQP-----SWNTRFPMIATTS-----EPHL 744
QY 4644 ELEVVAAEAGEVIMHKGMERIOFGG-----RFEVWSQGRQOMLVKFGTAEPDQGEYHCLGAQ 4699
Db 745 IVEDVTDITTLTKWRPPNRIKAGGIDGVLVEYCLEGSEEWV-----PANTEPVERCQFTV 799
QY 4700 GSICPAATFOVALSPASVDEAPQPSLPPAEQAQEGDLHLWEALARKRMSRETLDSIS 4759
Db 800 KNL-PTGARILFRVGVNIGRTEPA-----TLAQPVTTIREIA 836
QY 4760 ELPEDGRSQRLPOAEAEVAPDLSEGYSTADELARTGADLSHTSSDDESRAGTFSLVTY 4819
Db 837 EPPK-----IRLPHLRQ-----TY 851
QY 4820 LKKA-----GRGTSPLASKVGAAPSVKPOQOQOEPLAAVRPPLGLDUST----- 4864
Db 852 IRKVGELQNLNVFPFOCKERPQVWTKGAPLDTSS-----RVHVRTSDFTVFFV 900
QY 4865 -----KDLGDSMDKAAVKIOAFAKGYKVRKEMQOQEGPMFSHTFGDTEAQVGDALALEC 4919
Db 901 ROARSDSGEVEL-----SVQIENKMDTATIRVVEKAGPPIN-----VMVKE 944
QY 4920 VVASKADYRRLWKDGVLTGRRHHHIDQLGDGTCSLLIAGLDRADAGCYTCQVSNKFGQ 4979
Db 945 VWGTNALVEMQAPKD-----DGNSEIMGYFVQKADK--KTMENFNVYER 986
QY 4980 VTHSCAVVVGSESESESSGGLDDAFRAARLRLFRKSPAEVSDLELFLSADEGP 5039
Db 987 NRHTSCTV-----SDLIVGNEYFF----- 1005

QY 5040 APEPSPADWOTYREDEHFICIRFEALTEARQAVTRFOEMFATLIGIGVIEKLVEQGPFRVE 5099
Db 1006 -----RVYTEN---IC-----GLSDSPGVSKNTARILKT---GIFKFEYKEHDFR 1046
QY 5100 MCISKETPAPVVPPEPLPSLLTSDAAPVFLTELQNCQEVQDGPVPSFDCVVTGQPMPSVRW 5159
Db 1047 M-----APKFLTPLIDRVVAVAGYSAALNCAVRGHPKPKVVM 1082
QY 5160 FKDGKLLBEDDHVMTNEDQOGHQLITAVVPADGVVYRCLAENSMG--VSTTKAELRV 5216
Db 1083 MKNKMEIREDPKFLIT-NYQGVLTUNIRRPFFDAGTTCRAVNELGEALAECKLEVRV 1140
RESULT 15
ID UN52 CAEEL STANDARD; PRT: 3375 AA.
AC Q06561; O18263; Q9XTD2; Q9XTI5;
DT 01-JUN-1994 (Rel. 29, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane proteoglycan precursor (Perlecan homolog)
DE (Uncoordinated protein 52).
GN UNC-52 OR ZC101.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
[1]
RN SEQUENCE FROM N.A. (ISOFORM A), AND FUNCTION.
RX MEDLINE=93339574; PubMed=8393416;
RA Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.;
RT "Products of the unc-52 gene in Caenorhabditis elegans are homologous
RT to the core protein of the mammalian basement membrane heparan
RT sulfate proteoglycan.";
RL Genes Dev. 7:1471-1484(1993).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Percy C.M., Baynes C.;
RN Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN REVISIONS, AND ALTERNATIVE SPLICING.
RP Durbin R.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RL -/- FUNCTION: Probable role in myofilament assembly and/or attachment
CC of the myofilament lattice to the cell membrane. May be an
CC extracellular anchor for integrin receptors in muscle.
CC -/- SUBCELLULAR LOCATION: Extracellular matrix.
CC -/- ALTERNATIVE PRODUCTS:
CC Event=alternative splicing; Named isoforms=4;
CC Name=e;
CC IsoId=Q06561-1; Sequence=Displayed;
CC Note=No experimental confirmation available;
CC Name=a;
CC IsoId=Q06561-2; Sequence=VSP_007195, VSP_007196;
CC Name=b;
CC IsoId=Q06561-3; Sequence=VSP_007191, VSP_007192;
CC Note=No experimental confirmation available;
CC Name=c;
CC IsoId=Q06561-4; Sequence=VSP_007193, VSP_007194, VSP_007195,
CC VSP_007196;
CC Note=No experimental confirmation available;
CC -/- TISSUE SPECIFICITY: Found in the basement membrane of all
CC contractile tissues. It is concentrated over muscle dense bodies
CC and M-lines which are associated with beta-integrin.
CC -/- DEVELOPMENTAL STAGE: Synthesized early in embryogenesis.
CC -/- SIMILARITY: Contains 3 LDL-receptor class A domains.
CC -/- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.
CC -/- SIMILARITY: Contains 7 laminin EGF-like domains.
CC -/- SIMILARITY: Contains 3 laminin G-like domains.
CC -/- SIMILARITY: Contains 2 laminin IV domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----

DR EMBL; L13458; AAB28156.1; -
 DR EMBL; Z93375; CAB07567.1; -
 DR EMBL; Z93375; CAB07567.1; JOINED.
 DR EMBL; Z93375; CAB07568.1; -
 DR EMBL; Z93375; CAB07568.1; JOINED.
 DR EMBL; Z93375; CAB07569.1; -
 DR EMBL; Z93375; CAB07569.1; JOINED.
 DR EMBL; Z93375; CAB07704.1; -
 DR EMBL; Z93375; CAB07704.1; -
 DR EMBL; Z93375; CAB07706.1; -
 DR EMBL; Z93375; CAB07706.1; JOINED.
 DR EMBL; Z93375; CAB07707.1; -
 DR EMBL; Z93375; CAB07707.1; JOINED.
 DR EMBL; Z93375; CAB07708.1; -
 DR EMBL; Z93375; CAB07708.1; JOINED.
 DR HSP; P01130; 1LDR.
 DR WormPep; ZC101.2a; CE15028.
 DR WormPep; ZC101.2b; CE15030.
 DR WormPep; ZC101.2c; CE15034.
 DR WormPep; ZC101.2e; CE18424.
 DR GO; GO:0030239; P:myofibrillar assembly; IEP.
 DR GO; GO:0005578; C:extracellular matrix; IEP.
 DR InterPro; IP000895; ConA_like_tec_sl.
 DR InterPro; IP001881; EGF_Ca.
 DR InterPro; IP0006209; EGF_like.
 DR InterPro; IP0006210; EGF.
 DR InterPro; IP0007110; Ig-like.
 DR InterPro; IP0003599; Ig.
 DR InterPro; IP0003598; Ig_c2.
 DR InterPro; IP0000034; Laminin_B.
 DR InterPro; IP0002049; Laminin_EGF.
 DR InterPro; IP001791; Laminin_G.
 DR InterPro; IP0002172; LDL_receptor_A.
 DR Pfam; PF00047; Ig; 16.
 DR Pfam; PF00052; laminin_B; 2.
 DR Pfam; PF00053; laminin_EGF; 5.
 DR Pfam; PF00057; ldl_recept_a; 3.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR ProDom; PD003031; Laminin_B; 2.
 DR SMART; SM00181; EGF; 6.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00180; EGF_Lam; 6.
 DR SMART; SM00409; Ig; 17.
 DR SMART; SM00408; Ig_c2; 17.
 DR SMART; SM00381; Lamb; 1.
 DR SMART; SM00282; Lamb; 3.
 DR SMART; SM00192; LDLa; 3.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS00026; EGF_3; 3.
 DR PROSITE; PS00835; Ig_Like; 17.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
 DR PROSITE; PS01209; LDLRA_1; 3.
 DR PROSITE; PS00068; LDLRA_2; 3.
 KW Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;
 KW Basement membrane; Extracellular matrix; Alternative splicing;
 KW Laminin EGF-like domain.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 3375 BASEMENT MEMBRANE PROTEOGLYCAN.
 FT DOMAIN 45 130 IG-LIKE C2-TYPE 1.
 FT DOMAIN 148 184 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 189 225 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 232 269 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 271 355 IG-LIKE C2-TYPE 2.
 FT DOMAIN 384 431 LAMININ EGF-LIKE 1 (INCOMPLETE).
 FT DOMAIN 432 441 LAMININ EGF-LIKE 2 (N-TERMINAL).

FT DOMAIN 442 633 LAMININ DOMAIN IV 1.
 FT DOMAIN 634 666 LAMININ EGF-LIKE 2 (C-TERMINAL).
 FT DOMAIN 674 720 LAMININ EGF-LIKE 3 (INCOMPLETE).
 FT DOMAIN 721 730 LAMININ EGF-LIKE 4 (N-TERMINAL).
 FT DOMAIN 731 921 LAMININ DOMAIN IV 2.
 FT DOMAIN 922 954 LAMININ EGF-LIKE 4 (C-TERMINAL).
 FT DOMAIN 955 1004 LAMININ EGF-LIKE 5.
 FT DOMAIN 1011 1060 LAMININ EGF-LIKE 6.
 FT DOMAIN 1061 1111 LAMININ EGF-LIKE 7.
 FT DOMAIN 1126 1222 IG-LIKE C2-TYPE 3.
 FT DOMAIN 1226 1311 IG-LIKE C2-TYPE 4.
 FT DOMAIN 1319 1401 IG-LIKE C2-TYPE 5.
 FT DOMAIN 1410 1499 IG-LIKE C2-TYPE 6.
 FT DOMAIN 1503 1585 IG-LIKE C2-TYPE 7.
 FT DOMAIN 1588 1800 IG-LIKE C2-TYPE 8.
 FT DOMAIN 1690 1785 IG-LIKE C2-TYPE 9.
 FT DOMAIN 1793 1878 IG-LIKE C2-TYPE 10.
 FT DOMAIN 1886 1970 IG-LIKE C2-TYPE 11.
 FT DOMAIN 1973 2069 IG-LIKE C2-TYPE 12.
 FT DOMAIN 2073 2163 IG-LIKE C2-TYPE 13.
 FT DOMAIN 2173 2260 IG-LIKE C2-TYPE 14.
 FT DOMAIN 2263 2343 IG-LIKE C2-TYPE 15.
 FT DOMAIN 2349 2435 IG-LIKE C2-TYPE 16.
 FT DOMAIN 2446 2530 IG-LIKE C2-TYPE 17.
 FT DOMAIN 2532 2713 LAMININ G-LIKE 1.
 FT DOMAIN 2793 2960 LAMININ G-LIKE 2.
 FT DOMAIN 2961 3093 GLU-RICH.
 FT DOMAIN 2972 3066 THR-RICH.
 FT DOMAIN 3180 LAMININ G-LIKE 3.
 FT DISULFID 66 114 BY SIMILARITY.
 FT DISULFID 149 161 BY SIMILARITY.
 FT DISULFID 156 174 BY SIMILARITY.
 FT DISULFID 168 183 BY SIMILARITY.
 FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 197 215 BY SIMILARITY.
 FT DISULFID 209 224 BY SIMILARITY.
 FT DISULFID 230 246 BY SIMILARITY.
 FT DISULFID 240 259 BY SIMILARITY.
 FT DISULFID 253 268 BY SIMILARITY.
 FT DISULFID 955 964 BY SIMILARITY.
 FT DISULFID 957 971 BY SIMILARITY.
 FT DISULFID 974 983 BY SIMILARITY.
 FT DISULFID 986 1002 BY SIMILARITY.
 FT DISULFID 1011 1021 BY SIMILARITY.
 FT DISULFID 1013 1027 BY SIMILARITY.
 FT DISULFID 1030 1039 BY SIMILARITY.
 FT DISULFID 1042 1058 BY SIMILARITY.
 FT DISULFID 1061 1069 BY SIMILARITY.
 FT DISULFID 1063 1079 BY SIMILARITY.
 FT DISULFID 1082 1091 BY SIMILARITY.
 FT DISULFID 1094 1109 BY SIMILARITY.
 FT DISULFID 1152 1200 BY SIMILARITY.
 FT DISULFID 1338 1384 BY SIMILARITY.
 FT DISULFID 1435 1481 BY SIMILARITY.
 FT DISULFID 1527 1573 BY SIMILARITY.
 FT DISULFID 1618 1663 BY SIMILARITY.
 FT DISULFID 1719 1767 BY SIMILARITY.
 FT DISULFID 1814 1861 BY SIMILARITY.
 FT DISULFID 1907 1954 BY SIMILARITY.
 FT DISULFID 2053 2099 BY SIMILARITY.
 FT DISULFID 2099 2147 BY SIMILARITY.
 FT DISULFID 2195 2242 BY SIMILARITY.
 FT DISULFID 2284 2329 BY SIMILARITY.
 FT DISULFID 2374 2420 BY SIMILARITY.
 FT DISULFID 2467 2514 BY SIMILARITY.
 FT CARBOHYD 1422 1422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2476 2476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2950 2950 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3143 3143 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 1.3%; Score 519; DB 1; Length 3375;
 Best Local Similarity 20.0%; Pred. No. 4.8e-10;
 Matches 709; Conservative 415; Mismatches 1321; Indels 1108; Gaps 185;


```
Db 1927 LPHDVTLEPNILDFSNR-SELNGDYTCTASNPIGEASDHGNVNI----- 1971
QY 1976 SSSAALRVREAPVLFKKLEP-----QTVBERSSTVILEV--ELTRPWPPELRWTRNATALA 2028
Db 1972 -----GPSLTVKTNPPGPKLIVTGEPLQVKEAFCAPOCPPEPEVWENLHD----- 2016
QY 2029 PG-----KNVEIHAEGBARHLVHNVPFADRGFFGCEPDDKTQAKLTVEMRQVR 2078
Db 2017 PGPBGRDLPDFKPVITSEQPIRH-----PNVGLGNAGVYTKGSSAHATATKMIYIEVVE 2072
QY 2079 LVR-----GLQVAREQO-----TAT---MEVOLSHADVDGSMTRGLRFQOQPTC 2122
Db 2073 PSRTATVSVILGSSQWFDQEGKELICTATGSSLVDRLEWEKVDQDQLPDVEENEPGLL 2132
QY 2123 HL-----AVRGMHTLTLISGLRPDSGLMVFKAEVHTSARLVVVTLPVFSRPLQDVYT 2177
Db 2133 HFPSPKNSYAGEYR---CNGYR---NNEIIASAAVHVHSSAN--ADDEPKVEIEBPPVRVV 2185
QY 2178 TEKEKVTLECEL--SRPNVDVRM--LXQGV---ELRAGKTMATAAQACRSLTIVRCEFA 2230
Db 2186 SQGNIVLKCQVQAEHGEHFKWALLRGGSLVRQLGTEPTEITKADPPSNDPFGYRCNVE 2245
QY 2231 DQGVYVCDADHAQSSAVKVGRTYTLIYRVLAEADAGEIQFVAENAESRAQLRVKELPV 2290
Db 2246 DNNGLVIGS--AFTAVSVGQODKSHAQI---VKFDDKSDASFT----- 2283
QY 2291 TLVRPLRDKIAMKHGVLQCV--SRASAQRVHFQSOELQPGPKYELVSDGLYRKLIIIS 2349
Db 2284 -----CPIYVFGSKVDWYENGDLPS-----KAVPNG--NKIEYK 2317
QY 2350 DVHAEDDITYC-----DAGDVK--TSAQFFVEQSIIVRGLOQDVTVMEPA-----PAWFEC 2400
Db 2318 BFDDASAGTVYCKVSFGNVGVPVTAQMFVPD---TIIQVLEVSSESQIGDRAWFDC 2374
QY 2401 E-TSIPV-----RPPKWLKTKVLQAGNVGL-----EQEGTVHRLMLRRTCTST 2444
Db 2375 KVTGDSAVISWTKEGNDLPP-----NAQVTGGRLFTDLKEDNAGVYRCVAK---T 2424
QY 2445 MTGPVHF-TV-----GKSRSSARLWSDIPVWLTRPLEPKTGRELOSVLSCDFRPAPK- 2497
Db 2425 KAGPLQVETVLNVGSGQDQVTFVADSLPVVT-----VG---QPAYLSCICKTETXP 2475
QY 2498 --AVQWYKDDTPLSPSEKFKXGMAELRIURLMPADAGVYRCQAGSAHSSTEVTVEA 2555
Db 2476 NQSVVWTKKEGDL-PS-----GSRVEQGVLM-----PSVHRDDE----- 2509
QY 2556 REVTVTGLQDAEATEEGMASFCSELSHEDVEEWSLNGMPLYNDSFHEISHKGRHTLV 2615
Db 2510 -----GSYCEIVKEENPV---FSTVDLQIDDFIPV----- 2537
QY 2616 LKSIQRADAGIVRASSLKVSTARSLEVRVYVFLKALDLSAEBRTIALQ--CEVSDP 2673
Db 2538 -----IDGEP1-----ELPPLSDEEIVNLDEITLNTANP 2567
QY 2674 EAHVVRKDGVOQLP--SDKYDFLHTAGTGLVHVDSVPDAG--LYTCHVGS-----EET 2725
Db 2568 KG-IIFETKRINSGLLATPYDTH-----HEAKTDYGTVLYEFDIGNGROIET 2617
QY 2726 RARVRHDLHVGTITKRLKTMVLEGESCSPECVLSHESAD-----PAMTVVGKTVG 2778
Db 2618 TNPINPNEWNVIKNDKNQVTIQ-----LNDESATIRQHTNPLPSLSTGVNRPVF 2668
QY 2779 SSSRFQATROGRKVLIVVREAPSDAGEVVSFVSGTASLIVRERPAALIKPLEDQWV 2838
Db 2669 IGRHEPTNEANDF-----RGIIIS-----QYV 2690
QY 2839 APGEDVELRCELRAGTFPHWLKDKRAIKRSQKYDVVCEGTAMLVIRGASLKDA----- 2893
Db 2691 LSGHNV-----GLGDARIPSSVWXYD-ACAST--NLCLNGANCENANNHNG 2733
QY 2894 -----GEY-----TCEVEASKSTASLHVEBKANCFTB-ELTNLOVEEKGTA-- 2934
```

```
Db 2734 PSCECAEBFHBGEYQWMRNSNSCHDESCNTGICLDNEESQWCVPLGTTGLRCBEK-TELPQ 2792
QY 2935 ----FTCKTEHPAA--TVTWKGLLELRA---SGKH-----OPSQ 2965
Db 2793 PLGFTSDTSFLAVKRPVKFESIKMKLRPQADSDEHILMYFASDYGSNTKQVTSLSLIANQ 2852
QY 2966 EGLTLRLTISALEKADSDTYTCD-----ICQAQSRQALLVQGRVHIIEDLEDVDVQEG 3019
Db 2853 VVLTVRRPDKEVKIRSETLEAGELIDVAVRQAGNALVMTVDGNQVSTIE---TDTLKP 2908
QY 3020 SSATERCRISPANYEPVHWFLDKTPHLANELNEIDAQPGGYHVLTLRO---LALKDSGT 3075
Db 2909 GTEIFIGLPGCLNSP-----DDVVEQSFGCVVEIILNSQVDVLQNLSSGD 2956
QY 3076 IY-----FEAGQORASAAIR---VTEKPSVFSRELTDTATE--GEDTLVCTETST 3121
Db 2957 ISSCEESQFPVEEDDTTITTTTTEPPEAVIERPTTEPTTTEP.TEPTTEPTTTEPTT 3016
QY 3122 CDIPMCWTKDGKT 3134
Db 3017 TEEPTTTTEPTT 3029
```

Search completed: September 13, 2004, 11:25:53
Job time : 127 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:12:23 ; Search time 461 Seconds
(without alignments)
5453.469 Million cell updates/sec

Title: US-10-077-130-5
Perfect score: 41273
Sequence: 1 MDQPSGAPFLTRPKAFV.....RNREKRALLYKHNLAQVR 7968

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------|--------------------|
| 1 | 32134 | 77.9 | 6620 | Q96AA2 | Q96aa2 homo sapien |
| 2 | 8486.5 | 20.6 | 2212 | Q8NH3 | Q8nhn3 homo sapien |
| 3 | 6302 | 16.7 | 1319 | Q8HCD3 | Q8hcd3 homo sapien |
| 4 | 6619.5 | 16.0 | 1596 | Q8HCL6 | Q8hcl6 homo sapien |
| 5 | 5358.5 | 13.0 | 1040 | Q8NH2 | Q8nhn2 homo sapien |
| 6 | 3865 | 9.4 | 767 | Q8GYC7 | Q8gyc7 homo sapien |
| 7 | 3580.5 | 8.7 | 26926 | Q8WZB3 | Q8wzb3 homo sapien |
| 8 | 3577.5 | 8.7 | 26926 | Q10466 | Q10466 homo sapien |
| 9 | 3427 | 8.3 | 34350 | Q8WZ42 | Q8wz42 homo sapien |
| 10 | 3389 | 8.2 | 658 | Q8NH4 | Q8nhn4 homo sapien |
| 11 | 3313 | 8.0 | 7962 | Q10465 | Q10465 homo sapien |
| 12 | 3270 | 7.9 | 1020 | Q8NH0 | Q8nhn0 homo sapien |
| 13 | 3182.5 | 7.7 | 8081 | Q7Z120 | Q7z120 caenorhabdi |
| 14 | 3113 | 7.5 | 646 | Q8NH8 | Q8nhn8 homo sapien |
| 15 | 2793 | 6.8 | 551 | Q8NH7 | Q8nhn7 homo sapien |
| 16 | 2653 | 6.4 | 542 | Q8NH5 | Q8nhn5 homo sapien |

Q98918 gallus galli
Q7n00 rattus norv
Q7zz46 brachydanio
Q7zz48 brachydanio
Q9eqj5 mus musculu
Q8wz53 homo sapien
Q7yrf5 canis famil
Q9nfs3 drosophila
Q917u4 drosophila
Q96rw7 homo sapien
Q9b2p9 homo sapien
Q76518 caenorhabdi
Q81013 caenorhabdi
Q80zfs rattus norv
Q95ym2 procambarus
Q81103 bombyx mori
Q8sgd6 procambarus
Q95ym1 procambarus
Q9w055 drosophila
Q9n188 drosophila
Q9v4f7 drosophila
Q8isf6 caenorhabdi
Q8isf7 caenorhabdi
Q23551 caenorhabdi
Q8mld9 drosophila
Q801w8 brachydanio
Q23550 caenorhabdi
Q811h7 mus musculu
Q7yt99 mytilus gal

17 2629 6.4 4162 13 Q98918
18 2392.5 5.8 4076 11 Q7N00
19 2291 5.6 4039 13 Q7ZZ46
20 2273 5.5 5516 13 Q7ZZ48
21 2203 5.3 3262 11 Q9EQJ5
22 2189.5 5.3 5604 4 Q8WZ53
23 2168 5.3 3950 6 Q7YRF5
24 2087 5.1 16215 5 Q9NFS3
25 2087 5.1 18074 5 Q917U4
26 2072.5 5.0 5636 4 Q96RW7
27 2067 5.0 2242 4 Q9B2P9
28 2048 5.0 5198 5 Q76518
29 2043.5 5.0 5175 5 Q810L3
30 1996.5 4.8 803 11 Q80ZF5
31 1956.5 4.7 17352 5 Q95YM2
32 1896 4.6 4816 5 Q81103
33 1881.5 4.6 8625 5 Q8SGD6
34 1837 4.5 4824 5 Q95YM1
35 1763.5 4.3 4796 5 Q9W055
36 1736.5 4.2 4796 5 Q9N188
37 1677 4.1 8943 5 Q9V4F7
38 1665 4.0 18519 5 Q8ISF6
39 1664.5 4.0 18534 5 Q8ISF7
40 1650.5 4.0 7158 5 Q23551
41 1635 4.0 9270 5 Q8MLD9
42 1619.5 3.9 19066 13 Q801W8
43 1619 3.9 6831 5 Q23550
44 1612 3.9 345 11 Q811H7
45 1557.5 3.8 4736 5 Q7YT99

ALIGNMENTS

RESULT 1

Q96AA2 PRELIMINARY; PRT; 6620 AA.
ID Q96AA2
AC Q96AA2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Obscurin.
GN OBSCN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=21342081; PubMed=11448995;
RA Young P.W., Ehler E., Gautel M.;
RT "Obscurin, a giant sarcomeric rho guanine nucleotide exchange factor
protein involved in sarcomere assembly.";
RL J. Cell Biol. 154:123-136(2001).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
DR EMBL; AJ002535; CAC44768.1; -.
DR Genew; HGNC:15719; OBSCN.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004812; F:RNA ligase activity; IEA.
DR GO; GO:0006418; P:amino acid activation; IEA.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR000048; IQ-region.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00041; In3; 2.
DR Pfam; PF00047; Ig; 49.
DR Pfam; PF00612; IQ; 1.

| | | | | | |
|--|---|----|------|--|------|
| DR | Pfam: PF00169; PH; 1. | DR | 781 | RGRIAYQLSVQGLARFLHKDMAGSCVDAVAGGPAQFECETSEAHVHVHVDKGMELGHS | 840 |
| DR | Pfam: PF00621; RHOGEF; 1. | DR | 841 | GERFLQEDVGTTRRLVAATVTTRQDEGTYSRCRVGDSVDRLRVSEPKVVFVFAKEQLARRKL | 900 |
| DR | SMART; SM00060; FN3; 2. | DR | 841 | GERFLQEDVGTTRRLVAATVTTRQDEGTYSRCRVGDSVDRLRVSEPKVVFVFAKEQLARRKL | 900 |
| DR | SMART; SM00408; IGC2; 17. | DR | 901 | QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVCMNEATGCTTRRLVVCQAGQADAGEY | 960 |
| DR | SMART; SM00015; IQ; 1. | DR | 901 | QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVCMNEATGCTTRRLVVCQAGQADAGEY | 960 |
| DR | SMART; SM00233; PH; 1. | DR | 961 | SCEAGGQRLSPFLDVKEPKVVFVFAKQVVAHSEVQAEAGANATLSCEVAQAQAEVVMYKDGK | 1020 |
| DR | SMART; SM00326; SH3; 1. | DR | 961 | SCEAGGQRLSPFLDVKEPKVVFVFAKQVVAHSEVQAEAGANATLSCEVAQAQAEVVMYKDGK | 1020 |
| DR | PROSITE; PS00178; AA_TRNA_LIGASE_I; 1. | DR | 1021 | KLSSSLKLVHVEAKGCRRLRVQOAGKTADAGDYSCEARGQVSRFLHITTEPKMFAKEQSV | 1080 |
| DR | PROSITE; PS001010; DH_2; 1. | DR | 1021 | KLSSSLKLVHVEAKGCRRLRVQOAGKTADAGDYSCEARGQVSRFLHITTEPKMFAKEQSV | 1080 |
| DR | PROSITE; PS50835; IG_LIKE; 46. | DR | 1081 | HNEVOAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSSKVGMVEVKGCTRRRLVLPQAGKAD | 1140 |
| DR | PROSITE; PS50096; IQ; 1. | DR | 1081 | HNEVOAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSSKVGMVEVKGCTRRRLVLPQAGKAD | 1140 |
| DR | PROSITE; PS50003; PH_DOMAIN; 1. | DR | 1141 | AGEYSCGAGGQVSRFLHITTEPKVFAKEQSVHNEVOAEAGTATTAMLSCVAQAQTEVTWY | 1200 |
| KW | Immunoglobulin domain. | DR | 1141 | AGEYSCGAGGQVSRFLHITTEPKVFAKEQSVHNEVOAEAGTATTAMLSCVAQAQTEVTWY | 1200 |
| SQ | SEQUENCE 6620 AA; 721665 MW; C2AE8EB77B284452 CRC64; | DR | 1201 | KDGKLSKSSSKVGMVEVKGCTRRRLVVCQAGQADAGEYSCGAGGQVSRFLHITTEPKVFAK | 1260 |
| Query Match | 77.9%; Score 32134; DB 4; Length 6620; | DR | 1201 | KDGKLSKSSSKVGMVEVKGCTRRRLVVCQAGQADAGEYSCGAGGQVSRFLHITTEPKVFAK | 1260 |
| Best Local Similarity | 98.1%; Pred. No. 0; | DR | 1261 | EOLVHNEVTRTAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRIEAAAGCMRLVVCQA | 1320 |
| Matches 6250; Conservative 15; Mismatches 57; Indels 48; Gaps 9; | | DR | 1261 | EOLVHNEVTRTAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRIEAAAGCMRLVVCQA | 1320 |
| QY | 1 MDQPFSGAPFLTRPKAFVSVSGKDATLSQIVGNFTPOVSWKDDQOPVTAGARFLAQ | QY | 1321 | GOADAGEYTCGAGGQRLSFHLDVSEPKVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQVFAK | 1380 |
| DB | 1 MDQPFSGAPFLTRPKAFVSVSGKDATLSQIVGNFTPOVSWKDDQOPVTAGARFLAQ | DB | 1321 | GOADAGEYTCGAGGQRLSFHLDVSEPKVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQVFAK | 1380 |
| QY | 61 DGDLYRLTILDALGDSQYVCRNAINAIGFAFVAVGLQVDAEACAFQAFHLLRPTSR | QY | 1381 | VTWKDGKLSKSSSKVMEAVGCTRRRLVVCQACQADTGEYSCGAGGQRLSFSLDVAPKV | 1440 |
| DB | 61 DGDLYRLTILDALGDSQYVCRNAINAIGFAFVAVGLQVDAEACAFQAFHLLRPTSR | DB | 1381 | VTWKDGKLSKSSSKVMEAVGCTRRRLVVCQACQADTGEYSCGAGGQRLSFSLDVAPKV | 1440 |
| QY | 121 VREGSEATFCRCVGSPPAVSWKDGRRLLGEPDPRVVEELGEASALRIARPRDGG | QY | 1441 | VFAKEQVHREVOAQAGASTTSLCEVAQAQTEVTWYKDGKLSKSSSKVMEAVGCTRRLV | 1500 |
| DB | 121 VREGSEATFCRCVGSPPAVSWKDGRRLLGEPDPRVVEELGEASALRIARPRDGG | DB | 1441 | VFAKEQVHREVOAQAGASTTSLCEVAQAQTEVTWYKDGKLSKSSSKVMEAVGCTRRLV | 1500 |
| QY | 181 TYEVAENPLGAASAAAALVVDSDAATASRPGTSTAALLAHQRRREAVRAEGAPASPP | QY | 1501 | VQAGQADAGEYSCGAGGQRLSFHLDVSEPKVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQV | 1560 |
| DB | 181 TYEVAENPLGAASAAAALVVDSDAATASRPGTSTAALLAHQRRREAVRAEGAPASPP | DB | 1501 | VQAGQADAGEYSCGAGGQRLSFHLDVSEPKVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQV | 1560 |
| QY | 241 STGTRCTVTEGKHARLSQVYTGPKPETWKKDGLQVTEGRRHVVDACENFVLKILF | QY | 1561 | AQTEVTWYKDGKLSKSSSKVMEAVGCTRRRLVVCQAGQADAGEYSCGAGGQRLSFHLDV | 1620 |
| DB | 241 STGTRCTVTEGKHARLSQVYTGPKPETWKKDGLQVTEGRRHVVDACENFVLKILF | DB | 1561 | AQTEVTWYKDGKLSKSSSKVMEAVGCTRRRLVVCQAGQADAGEYSCGAGGQRLSFHLDV | 1620 |
| QY | 301 CKQSDRGLYTCASNLYGQYSSVYVVRPAVPFKKRLQDLVREKESATFLCEVPQPS | QY | 1621 | EPKVVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQV | 1680 |
| DB | 301 CKQSDRGLYTCASNLYGQYSSVYVVRPAVPFKKRLQDLVREKESATFLCEVPQPS | DB | 1621 | EPKVVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQVFAKEQV | 1680 |
| QY | 361 TEAAWFKEETRLWASAKYIEEGTERRLLTVRNVSADDDAVYICETPEGSRTVAELAVQG | QY | 1681 | RRLVVCQAGQADAGEYSCGAGGQRLSFHLDVSEPKVFAKEQVFAKEQVFAKEQVFAKEQV | 1740 |
| DB | 361 TEAAWFKEETRLWASAKYIEEGTERRLLTVRNVSADDDAVYICETPEGSRTVAELAVQG | DB | 1681 | RRLVVCQAGQADAGEYSCGAGGQRLSFHLDVSEPKVFAKEQVFAKEQVFAKEQVFAKEQV | 1740 |
| QY | 421 NLLRKLPRKTAVRVGDVAMFCVELAVPVPVHVRNQEEVYVAGRVVAISAEGRHTLTIS | QY | 1741 | TLATPSAATVTLKDGVEIRRSKHEHETASQDHTLTVHGAQVLDLSAIYSCRVAEGQDF | 1800 |
| DB | 421 NLLRKLPRKTAVRVGDVAMFCVELAVPVPVHVRNQEEVYVAGRVVAISAEGRHTLTIS | DB | 1741 | TLATPSAATVTLKDGVEIRRSKHEHETASQDHTLTVHGAQVLDLSAIYSCRVAEGQDF | 1800 |
| QY | 481 QCCLEDVGVAFMAGDCQSTTRFCVSAAPRPDPVPPVDMVVAEENFQVSAVFNQSPY | QY | 1801 | PVQVEEVAAKFCRLLEPVCGLGCTVTLACELSPACAEVVMVRCGNTOPRVGKRFQVMAEG | 1860 |
| DB | 481 QCCLEDVGVAFMAGDCQSTTRFCVSAAPRPDPVPPVDMVVAEENFQVSAVFNQSPY | DB | 1801 | PVQVEEVAAKFCRLLEPVCGLGCTVTLACELSPACAEVVMVRCGNTOPRVGKRFQVMAEG | 1860 |
| QY | 541 RPVTIDGVLVEKKLGTWYTWIRCHEAEWATPELTVDVAEENFQVSAVFNQSPY | QY | 1861 | PVRSLTVLGRAEADAGEVCESDHHTSAQITVSVPRVVKFMSGLSTVVAEAGEATFQC | 1920 |
| DB | 541 RPVTIDGVLVEKKLGTWYTWIRCHEAEWATPELTVDVAEENFQVSAVFNQSPY | DB | 1861 | PVRSLTVLGRAEADAGEVCESDHHTSAQITVSVPRVVKFMSGLSTVVAEAGEATFQC | 1920 |
| QY | 601 LEPPGTVHLAPKLAVRPLKAVQAEVGEVTFVSDLTVAAGWFLDQALKASSVYEIH | QY | | | |
| DB | 601 LEPPGTVHLAPKLAVRPLKAVQAEVGEVTFVSDLTVAAGWFLDQALKASSVYEIH | DB | | | |
| QY | 661 CDRTRHTLTIREVPASLHQAQLKVFANGIESSIRMEVRAAPGLTANKPAAAREVLAEL | QY | | | |
| DB | 661 CDRTRHTLTIREVPASLHQAQLKVFANGIESSIRMEVRAAPGLTANKPAAAREVLAEL | DB | | | |
| QY | 721 HEAAQLAEISDQAAAVTWLKDGRITLSPGPKYEVQASAGRVLLVRDVARDDAGLYECVS | QY | | | |
| DB | 721 HEAAQLAEISDQAAAVTWLKDGRITLSPGPKYEVQASAGRVLLVRDVARDDAGLYECVS | DB | | | |
| QY | 781 RGGRIAYQLSVQGLARFLHKDMAGSCVDAVAGGPAQFECETSEAHVHVHVDKGMELGHS | QY | | | |

QY 1921 VVSPSDVAVVFRDGLALQSEKFAISQSGASHSLT:SDLVLEADAGQITVEAEGASSAA 1980
DB 1921 VVSPSDVAVVFRDGLALQSEKFAISQSGASHSLT:SDLVLEADAGQITVEAEGASSAA 1980
QY 1981 LRVRAPVLPFKKLEPOTVEERSSVTLEVLTRPWPBLRTRNATALAPGNVETIHAEGA 2040
DB 1981 LRVRAPVLPFKKLEPOTVEERSSVTLEVLTRPWPBLRTRNATALAPGNVETIHAEGA 2040
QY 2041 RHRLVHNHVGADRGFFCETPDDKTQAKLVEMRQVRLVRGLQAVAREQGTATMEVQL 2100
DB 2041 RHRLVHNHVGADRGFFCETPDDKTQAKLVEMRQVRLVRGLQAVAREQGTATMEVQL 2100
QY 2101 SHADVGSNTDGLRFQOGPTCHLAVRGPMTLTLISGLRPEDSGLMVFKABGVHTSARLV 2160
DB 2101 SHADVGSNTDGLRFQOGPTCHLAVRGPMTLTLISGLRPEDSGLMVFKABGVHTSARLV 2160
QY 2161 VTLPVSPSRPLQDVVTTTEKEKVTLECELSRPNVDVRLKDGVELRAGKTMIAAQAQACR 2220
DB 2161 VTLPVSPSRPLQDVVTTTEKEKVTLECELSRPNVDVRLKDGVELRAGKTMIAAQAQACR 2220
QY 2221 SLTYRCFADQGVVCDHAQAQSSASVKVQGRYTYLLYRVLAEADAGEIQFVAENABSR 2280
DB 2221 SLTYRCFADQGVVCDHAQAQSSASVKVQGRYTYLLYRVLAEADAGEIQFVAENABSR 2280
QY 2281 AOLRVKELPVTLRPLRDKIAMEXHGVLECOVSRASAOVRWFVXGSOBLQPGPKYELVSD 2340
DB 2281 AOLRVKELPVTLRPLRDKIAMEXHGVLECOVSRASAOVRWFVXGSOBLQPGPKYELVSD 2340
QY 2341 GLYRKLIIISDVHAEDEDYTCDAQGVKTSQAQFVEEQSITIVRGLQDVTVMEPAPAMFEC 2400
DB 2341 GLYRKLIIISDVHAEDEDYTCDAQGVKTSQAQFVEEQSITIVRGLQDVTVMEPAPAMFEC 2400
QY 2401 ETSIPSVRPPKWLIGKTVLQAGNVGLBOEGTVHRLMLRRTCSMTGPHVFTVCKSRSSA 2460
DB 2401 ETSIPSVRPPKWLIGKTVLQAGNVGLBOEGTVHRLMLRRTCSMTGPHVFTVCKSRSSA 2460
QY 2461 RLWVSDIPVLTTRPLEPTKRELQSVLSCDFRPAKAVQWYKDDTPLSPSEKFKMSLEG 2520
DB 2461 RLWVSDIPVLTTRPLEPTKRELQSVLSCDFRPAKAVQWYKDDTPLSPSEKFKMSLEG 2520
QY 2521 QMAELRIILRMPADAGVTRCOAGSAHSSTEVTVAREVTVTGPIQDAEABEGWASFSCE 2580
DB 2521 QMAELRIILRMPADAGVTRCOAGSAHSSTEVTVAREVTVTGPIQDAEABEGWASFSCE 2580
QY 2581 LSHEDEEVWSLNGMPLNDSFHEISHKGRHTLVLSKIQRADAGIVRASSLKVSTEARL 2640
DB 2581 LSHEDEEVWSLNGMPLNDSFHEISHKGRHTLVLSKIQRADAGIVRASSLKVSTEARL 2640
QY 2641 EVVRKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRKXGQVQLGSPDKYDFLHTAGT 2700
DB 2641 EVVRKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRKXGQVQLGSPDKYDFLHTAGT 2700
QY 2701 RGLVVDVSPDAGLYTCHVGSEETRAVRVHDLHVGITKELKTMEVLEGESCSFECVLS 2760
DB 2701 RGLVVDVSPDAGLYTCHVGSEETRAVRVHDLHVGITKELKTMEVLEGESCSFECVLS 2760
QY 2761 HESASDPAMWTVGGKTVGSSSRFQATROGRKYLIVREAPSDADAGEVVFVSRGLTSKASL 2820
DB 2761 HESASDPAMWTVGGKTVGSSSRFQATROGRKYLIVREAPSDADAGEVVFVSRGLTSKASL 2820
QY 2821 IVREPAALIKPLEDQWAPGEDVELRCELSRAGTPVHVLKDRKAIKSKQYDVVVCBGTM 2880
DB 2821 IVREPAALIKPLEDQWAPGEDVELRCELSRAGTPVHVLKDRKAIKSKQYDVVVCBGTM 2880
QY 2881 AMLVIRGASLKDAGEYTCVEBAKSTASLHVEEKANCFTBELTNLQVEKGTAVFTCKTE 2940
DB 2881 AMLVIRGASLKDAGEYTCVEBAKSTASLHVEEKANCFTBELTNLQVEKGTAVFTCKTE 2940
QY 2941 HPAATVTRKGLLELRASGHQPSQEGTLRLTITSALEKADSDYTCDIGAQASRAQLLV 3000
DB 2941 HPAATVTRKGLLELRASGHQPSQEGTLRLTITSALEKADSDYTCDIGAQASRAQLLV 3000

QY 3001 QQRVHHIIEDELDVQVEGSSATFRCRISPNYEPVHWFLDKTPHLANELNEIDAOPGY 3060
DB 3001 QQRVHHIIEDELDVQVEGSSATFRCRISPNYEPVHWFLDKTPHLANELNEIDAOPGY 3060
QY 3061 HVLTLRQLAKDSGTIYFAGQORASAAALRVTEKESVSFRELTDATITEGDLTLVCETS 3120
DB 3061 HVLTLRQLAKDSGTIYFAGQORASAAALRVTEKESVSFRELTDATITEGDLTLVCETS 3120
QY 3121 TCDIPMCWTCKDRTLRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRV 3180
DB 3121 TCDIPMCWTCKDRTLRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRV 3180
QY 3181 HARPVFQBALKDLEVLGGAATLRCLSSVAAPVKWCYGNVLPDGRKXSLRQEGAMLE 3240
DB 3181 HARPVFQBALKDLEVLGGAATLRCLSSVAAPVKWCYGNVLPDGRKXSLRQEGAMLE 3240
QY 3241 LVVRNLRPQDSGRYSCSPQDQTSATLTVTALPAQFIGIKLRNKEATEGATATLRCELSKT 3300
DB 3241 LVVRNLRPQDSGRYSCSPQDQTSATLTVTALPAQFIGIKLRNKEATEGATATLRCELSKT 3300
QY 3301 APVEMKGETLDDGRYCLRODGMCELOIISGLAVMDAAEYSCVCGEBERTSASLTIRPM 3360
DB 3301 APVEMKGETLDDGRYCLRODGMCELOIISGLAVMDAAEYSCVCGEBERTSASLTIRPM 3360
QY 3361 PAHFIGELRHQBSIEGATATLRCELSKAAAPVEMKGRESLRDGRHSLRQDGAVALCELOIC 3420
DB 3361 PAHFIGELRHQBSIEGATATLRCELSKAAAPVEMKGRESLRDGRHSLRQDGAVALCELOIC 3420
QY 3421 GLAVADAGYSCVCGEBERTSAILTVKALPAKTEGLRNEEAVEGATAMLCESKVAPVE 3480
DB 3421 GLAVADAGYSCVCGEBERTSAILTVKALPAKTEGLRNEEAVEGATAMLCESKVAPVE 3480
QY 3481 WRKGPENLRDGRYILRQEGTRCELOICGLAVADAGEYLCVCGQERTSATLTIRALPARF 3540
DB 3481 WRKGPENLRDGRYILRQEGTRCELOICGLAVADAGEYLCVCGQERTSATLTIRALPARF 3540
QY 3541 IEDVKQNEAREGATAVLQCELSAAPVEMKGETLDDGRYSLRQDGTKCELOIIRGLAM 3600
DB 3541 IEDVKQNEAREGATAVLQCELSAAPVEMKGETLDDGRYSLRQDGTKCELOIIRGLAM 3600
QY 3601 ADTGEYSCVCGQERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLCELSKVAPVEMWK 3660
DB 3601 ADTGEYSCVCGQERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLCELSKVAPVEMWK 3660
QY 3661 HETLRDGRHSLRQDGAVALCELOIIRGLVAEDAGEYLCVCGKERTSAMLTVRAMPSKFIEGL 3720
DB 3661 HETLRDGRHSLRQDGAVALCELOIIRGLVAEDAGEYLCVCGKERTSAMLTVRAMPSKFIEGL 3720
QY 3721 RNEEATEGATATLWCELSKAAAPVEMKGETLDDGRHSLRQDGSCELOIIRGLAVVDAG 3780
DB 3721 RNEEATEGATATLWCELSKAAAPVEMKGETLDDGRHSLRQDGSCELOIIRGLAVVDAG 3780
QY 3781 EYSCVCGQERTSATLTVRALPARFIEDVKNQAREGATAVLQCELSKAAAPVEMKGETL 3840
DB 3781 EYSCVCGQERTSATLTVRALPARFIEDVKNQAREGATAVLQCELSKAAAPVEMKGETL 3840
QY 3841 RGGDRYSLRQDGTTRCELOIHGLSVADTGEYSCVCGQERTSATLTVRAPQVPREPLOSLO 3900
DB 3841 RGGDRYSLRQDGTTRCELOIHGLSVADTGEYSCVCGQERTSATLTVRAPQVPREPLOSLO 3900
QY 3901 ABEGSTATTQCELSSEPTATVWMSKGLQLOANGREPRLOQGTAEVLQDLQREDTGEYT 3960
DB 3901 ABEGSTATTQCELSSEPTATVWMSKGLQLOANGREPRLOQGTAEVLQDLQREDTGEYT 3960
QY 3961 CTCGSQATSATLTVTAAAPVRFLRELOHQVEDEGGTAHLCCELSRAGASVEMKSGSLQLP 4020
DB 3961 CTCGSQATSATLTVTAAAPVRFLRELOHQVEDEGGTAHLCCELSRAGASVEMKSGSLQLP 4020
QY 4021 CAKYQMVQGAALVVRGVEQEDAGDYTCDTGHTOSMASLSVRVPRPKFKTRQLQSLQEB 4080
DB 4021 CAKYQMVQGAALVVRGVEQEDAGDYTCDTGHTOSMASLSVRVPRPKFKTRQLQSLQEB 4080
QY 4081 TGDIALCCOLSDAESGAVVQMLKEGVHAGPKYEMRSGQATRELLIHLQEAKTGEYA 4140

| | | | | | | |
|------|----|--|------|----|--|------|
| 4081 | DB | TGDIARLCCQLSDAESGAVVOWLKEGVHLAGPKYEMRSQATRELLIHQLEAKDTGEYA | 4140 | DB | KDGKLLBEDDHTWINEOQGGHLLIITAVVPADMGVYRCLAESNGVSSSTKAELRVDLTS | 5220 |
| 4141 | QY | CVTGQKTAASLRVTEPEVTIVRGLVDAEVTADDEVEFSCEVSPAGATGVQWCLQGLPLQ | 4200 | QY | TYDYTAADATSESSYFSAQGYLSREOEGTSTTDEGQLPQVVEBELRLQVAPGTRLAKF | 5280 |
| 4141 | DB | CVTGQKTAASLRVTEPEVTIVRGLVDAEVTADDEVEFSCEVSPAGATGVQWCLQGLPLQ | 4200 | DB | TYDYTAADATSESSYFSAQGYLSREOEGTSTTDEGQLPQVVEBELRLQVAPGTRLAKF | 5280 |
| 4201 | QY | SNTEVAVRGRIRHTLRKGVTPEDAGTVSFHGNHASSAQLTVRAPEVTILEPQDVQ | 4260 | QY | QKVKGYPAAPRILYFKOGQPLTASAHIRMTGKILHTLEII SVTREDSGQAAVYTSNANG | 5340 |
| 4201 | DB | SNTEVAVRGRIRHTLRKGVTPEDAGTVSFHGNHASSAQLTVRAPEVTILEPQDVQ | 4260 | DB | QKVKGYPAAPRILYFKOGQPLTASAHIRMTGKILHTLEII SVTREDSGQAAVYTSNANG | 5340 |
| 4261 | QY | LSEGDASFQRLSRASQCEARWALGGVPLQANEMNDITVEQGTLLHLLTKHVKLTEDAGT | 4320 | QY | AAVSSARLLVRGPDPEEKPSADVHEQLVPPRMLERFTPKVKVKGSSITFSVKVEGRVPP | 5400 |
| 4261 | DB | LSEGDASFQRLSRASQCEARWALGGVPLQANEMNDITVEQGTLLHLLTKHVKLTEDAGT | 4320 | DB | AAVSSARLLVRGPDPEEKPSADVHEQLVPPRMLERFTPKVKVKGSSITFSVKVEGRVPP | 5400 |
| 4321 | QY | VSFHVTGCSSEAOQLYTAKNITVVRGLENVVALEGEALFECOLSQPEVAATHWLLDDEPV | 4380 | QY | TVHMLREAEAGVLMIGPDTPGYTVASSAQOHSVLVDVGRHOHTYTCIASNAAGQALC | 5460 |
| 4321 | DB | VSFHVTGCSSEAOQLYTAKNITVVRGLENVVALEGEALFECOLSQPEVAATHWLLDDEPV | 4380 | DB | TVHMLREAEAGVLMIGPDTPGYTVASSAQOHSVLVDVGRHOHTYTCIASNAAGQALC | 5460 |
| 4381 | QY | RTSENAEVVFFENGRLHLLLNLRPQDSCHVTFLAGDMVTSAPLTVRGMRLLEPLKN | 4440 | QY | SASLHVSGLPKVEQEKVKEALISTFLOGTTQALISAQGLETASFADLGQKKEPPLAKE | 5520 |
| 4381 | DB | RTSENAEVVFFENGRLHLLLNLRPQDSCHVTFLAGDMVTSAPLTVRGMRLLEPLKN | 4440 | DB | SASLHVSGLPKVEQEKVKEALISTFLOGTTQALISAQGLETASFADLGQKKEPPLAKE | 5520 |
| 4441 | QY | AAVRAGAARTCTILSEAVPGBASWYINGAAVOPDDSDMTVTADGSHQALLRSAPHH | 4500 | QY | ALGHLSLAEVGTEFEFLQKLTQITQITMVSATKTOAKLOVPGGDSDEDSKTPSASPRHGRS | 5580 |
| 4441 | DB | AAVRAGAARTCTILSEAVPGBASWYINGAAVOPDDSDMTVTADGSHQALLRSAPHH | 4500 | DB | ALGHLSLAEVGTEFEFLQKLTQITQITMVSATKTOAKLOVPGGDSDEDSKTPSASPRHGRS | 5580 |
| 4501 | QY | AGEVTFACRDVAVASARLTVLGLPDPEDAEVVAHSSHTVLSMAAPMSDGGGLCGYRVE | 4560 | QY | PSSSTQESSSESDGDARGEIFDIYVVTADYLPFGAEQDAITLREGQVVEVLDAAHPLRW | 5640 |
| 4501 | DB | AGEVTFACRDVAVASARLTVLGLPDPEDAEVVAHSSHTVLSMAAPMSDGGGLCGYRVE | 4560 | DB | PSSSTQESSSESDGDARGEIFDIYVVTADYLPFGAEQDAITLREGQVVEVLDAAHPLRW | 5640 |
| 4561 | QY | VKEGATQWRLCHELVGPECVVDGLAPGETYFRVAAVGPVGAEPVHLPTQVRLAEP | 4620 | QY | LVRTPTKSSPSRQGWSPAYLDRRLKLSPEWGAABAEPEFGEAVSEDEYKARLSVIOE | 5700 |
| 4561 | DB | VKEGATQWRLCHELVGPECVVDGLAPGETYFRVAAVGPVGAEPVHLPTQVRLAEP | 4620 | DB | LVRTPTKSSPSRQGWSPAYLDRRLKLSPEWGAABAEPEFGEAVSEDEYKARLSVIOE | 5700 |
| 4621 | QY | KPVPPQPSAPSRQVAGEDVSELEVVAAEAGEVIMHKGMERLOPGGRFEVSGRQOML | 4680 | QY | LLSSQCAFVEELOFQSHLOHLERCPHPIAVAGQKAVIFRNVRDITGRPHSSFLQELQ | 5760 |
| 4621 | DB | KPVPPQPSAPSRQVAGEDVSELEVVAAEAGEVIMHKGMERLOPGGRFEVSGRQOML | 4680 | DB | LLSSQCAFVEELOFQSHLOHLERCPHPIAVAGQKAVIFRNVRDITGRPHSSFLQELQ | 5760 |
| 4681 | QY | VIKGFTABDQGEYHCGLAQGSICPAATFQVALSPASVDEAPQPSLPPEAAQEGDLHLW | 4740 | QY | CDTDDVAVACFTKQAAFEQYLEFLVGRVQAESVWVSTAIQEFYKKAEEALLAGDPSQP | 5820 |
| 4681 | DB | VIKGFTABDQGEYHCGLAQGSICPAATFQVALSPASVDEAPQPSLPPEAAQEGDLHLW | 4740 | DB | CDTDDVAVACFTKQAAFEQYLEFLVGRVQAESVWVSTAIQEFYKKAEEALLAGDPSQP | 5820 |
| 4741 | QY | EALARKRMSREPTLDSITSELPEEDGRSRLPQABEVAPODLSEGYSTADELARTGDADL | 4800 | QY | PPPLQHYLEQPVVERVQYQALLKELIRNKARNQNCALLEQAVVVSALPQRAENKLHV | 5880 |
| 4741 | DB | EALARKRMSREPTLDSITSELPEEDGRSRLPQABEVAPODLSEGYSTADELARTGDADL | 4800 | DB | PPPLQHYLEQPVVERVQYQALLKELIRNKARNQNCALLEQAVVVSALPQRAENKLHV | 5880 |
| 4801 | QY | SHTSSDDBSRAGTPSLVTLKACRPGTSPILASKVGAAPAPSVKPOQOQEPFLAAVRPPLG | 4860 | QY | SLMENYPTGLEALGPRIQGHFI VMEGAPGARVPMKGNHNVFLFRNHLVICKPRRDSRT | 5940 |
| 4801 | DB | SHTSSDDBSRAGTPSLVTLKACRPGTSPILASKVGAAPAPSVKPOQOQEPFLAAVRPPLG | 4860 | DB | SLMENYPTGLEALGPRIQGHFI VMEGAPGARVPMKGNHNVFLFRNHLVICKPRRDSRT | 5940 |
| 4861 | QY | DLSTKDLGDPNDKAAVKIQAAFKGYKVRKEMKQOEGPMFSHTFGTEAQVGDALRLCV | 4920 | QY | DTVSVVFRNMKLSIDLDNDQVGGDDRAFEVYQEREDSVRKVLLQARTAIITKSSWKEIC | 6000 |
| 4861 | DB | DLSTKDLGDPNDKAAVKIQAAFKGYKVRKEMKQOEGPMFSHTFGTEAQVGDALRLCV | 4920 | DB | DTVSVVFRNMKLSIDLDNDQVGGDDRAFEVYQEREDSVRKVLLQARTAIITKSSWKEIC | 6000 |
| 4921 | QY | VASKADVAVARLKDQVELTDGRHHIDQLGDTGCSLLIAGLDRADAGCYTCQVSNKFGOV | 4980 | QY | GIQORLALPVWRPDPFEELADCTAEALGETVKLACRVTGTPKPVISWYKDGKAVQVDPHH | 6060 |
| 4921 | DB | VASKADVAVARLKDQVELTDGRHHIDQLGDTGCSLLIAGLDRADAGCYTCQVSNKFGOV | 4980 | DB | GIQORLALPVWRPDPFEELADCTAEALGETVKLACRVTGTPKPVISWYKDGKAVQVDPHH | 6060 |
| 4981 | QY | THSACVVVSGSESAESSGGELDDAFRAARLRHLFRTKSPAESVDEBELFLSADEGPA | 5040 | QY | LIIEPDGSCALILDSLTGVDSGQVCMCPAASAGNCSTLGLILVQVPPRFNFKVRASFPV | 6120 |
| 4981 | DB | THSACVVVSGSESAESSGGELDDAFRAARLRHLFRTKSPAESVDEBELFLSADEGPA | 5040 | DB | LIIEPDGSCALILDSLTGVDSGQVCMCPAASAGNCSTLGLILVQVPPRFNFKVRASFPV | 6120 |
| 5041 | QY | EPESPADQWTVREDEHFCIRFEALTEARQAVTRFQEMFATLIGVBEIKLVEOGPRVEM | 5100 | QY | EGEDAQFTCTIEGAPYPOIRWYKDCALITGNKFTQJSEPRSGLLVIVRAASKEDLGLY | 6180 |
| 5041 | DB | EPESPADQWTVREDEHFCIRFEALTEARQAVTRFQEMFATLIGVBEIKLVEOGPRVEM | 5100 | DB | EGEDAQFTCTIEGAPYPOIRWYKDCALITGNKFTQJSEPRSGLLVIVRAASKEDLGLY | 6180 |
| 5101 | QY | CISKETPAPVPPPEPLSLTSDAAPVFLTELQNEQVQDGYPSVFCVVTGQPMPSVRWF | 5160 | QY | ECELVNRLGSRASAEALRIQSMLOAQOCHREOLVAAVEDTTLER | 6232 |
| 5101 | DB | CISKETPAPVPPPEPLSLTSDAAPVFLTELQNEQVQDGYPSVFCVVTGQPMPSVRWF | 5160 | DB | ECELVNRLGSRASAEALRIQSMLOAQOCHREOLVAAVEDTTLER | 6232 |
| 5161 | QY | KDGKLLBEDDHTWINEOQGGHLLIITAVVPADMGVYRCLAESNGVSSSTKAELRVDLTS | 5220 | QY | SVLKXELLGPKAPGSTGLTGPCCPRG-----APAL-----QSTGQOP- 6271 | |
| | | | | DB | TVVKSPPCORR-SFSKSPSRSPRCSASPLRPGLLAPOLLVLPAGQAPRAEPAGQKPV 6299 | |

QY 6272 -PVTGTSAPVPRV-----POPLLHEGPE-----OEPAIAAAQEWTPVIRMEGA 6317
 Db 6300 VPTLYTEAHSAHSPALGSLSPQPKVWEETIEVRVKMGPGQVSPTE--VPRSSGH 6357
 QY 6318 AW--PGAGTG 6325
 Db 6358 LFTLPGATPG 6367

RESULT 2
 Q8NH3 PRELIMINARY; PRT; 2212 AA.

AC Q8NH3;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Obscurin (Fragment).
 GN OBSN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Young P., Ehler E., Gautel M.;
 RT "Obscurin, a giant sarcomeric Rho-GAP protein involved in sarcomere
 RT assembly.";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ314905; CAC85752.1; -;
 DR InterPro; IPR003599; IG.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003598; IG_C2.
 DR Pfam; PF00047; IG; 24.
 DR SMART; SMO0409; IG; 25.
 DR SMART; SMO0408; IGC2; 23.
 DR PROSITE; PS00835; IG Like; 22.
 KW Immunoglobulin domain.
 FT NON TER
 FT NON TER
 SQ SEQUENCE 2212 AA; 241980 MW; B6DFA263EB749465 CRC64;

Query Match 20.6%; Score 8486.5; DB 4; Length 2212;
 Best Local Similarity 76.5%; Pred. No. 0;
 Matches 1692; Conservative 1; Mismatches 2; Indels 517; Gaps 3;

QY 2734 LHVGITKELKTMVELEGESCFECVLSHESADPAMTVGKTVGSSRFOATQGRKYI 2793
 Db 1 LHVGITKELKTMVELEGESCFECVLSHESADPAMTVGKTVGSSRFOATQGRKYI 60

QY 2794 LVREAAPS DAGEVVFVSRGLTSKASLIVRRPAAI IKPLEDQWVAPGEDVELRCELSRA 2853
 Db 61 LVREAAPS DAGEVVFVSRGLTSKASLIVRRPAAI IKPLEDQWVAPGEDVELRCELSRA 120

QY 2854 GTPVHNLKDKRAIRKSKYQV DVEGCTMAMLVIRGASLKDAGEYTCVEEASKSTASLHVEE 2913
 Db 121 GTPVHNLKDKRAIRKSKYQV DVEGCTMAMLVIRGASLKDAGEYTCVEEASKSTASLHVEE 180

QY 2914 KANCFTEELTNLQVEEKGTAFTCTKTEHPAATVTVWKGLLELRASGKHQPSQEGTLRLT 2973
 Db 181 KANCFTEELTNLQVEEKGTAFTCTKTEHPAATVTVWKGLLELRASGKHQPSQEGTLRLT 240

QY 2974 ISALEKADSDTYTCDIGQAQSRALLVQ-GRVHHIEDLVVDVQEGSSATFRCRISPAN 3032
 Db 241 ISALEKADSDTYTCDIGQAQSRALLVQAGRRVHHIEDLVVDVQEGSSATFRCRISPAN 300

QY 3033 YEPVHWFLDKTPHANELNEIDAOPGGVHVLTLROLAKDSGTIYFEAGDORASALRV 3092
 Db 301 YEPVHWFLDKTPHANELNEIDAOPGGVHVLTLROLAKDSGTIYFEAGDORASALRV 360

QY 3093 EXPSPVFSRELDTATITEGEDTLVCTETCDIPFCMTWKDGKTLRGSARCOLSHEGHRAL 3152
 Db 361 EXPSPVFSRELDTATITEGEDTLVCTETCDIPFCMTWKDGKTLRGSARCOLSHEGHRAL 420

QY 3153 LITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFOEALKOLEVLEGGAATLRCVLSVA 3212
 Db 421 LITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFOEALKOLEVLEGGAATLRCVLSVA 480

QY 3213 APVKWCYGNVLRPGDKYSLROEGAMLELVVRNLRPQDSGRYSCSFGQDTTSATLVTAL 3272
 Db 481 APVKWCYGNVLRPGDKYSLROEGAMLELVVRNLRPQDSGRYSCSFGQDTTSATLVTAL 540

QY 3273 PAQFIGKLRNKEATEGATATLRCELSKTAPVWRKGSSETLRDGDYRCLRDGAMCELOIR 3332
 Db 541 PAQFIGKLRNKEATEGATATLRCELSKTAPVWRKGSSETLRDGDYRCLRDGAMCELOIR 600

QY 3333 GLAMVDAAEYSCVCGEERTSASLTIRPMPAHFIRHQRHESLEGATATLRCELSKAAPVE 3392
 Db 601 GLAMVDAAEYSCVCGEERTSASLTIRPMPAHFIRHQRHESLEGATATLRCELSKAAPVE 660

QY 3393 WRKGRESLRDGDHSLRDGAVCELOICGLAVADAGEYSCVCGEERTSATLTVKALPAKF 3452
 Db 661 WRKGRESLRDGDHSLRDGAVCELOICGLAVADAGEYSCVCGEERTSATLTVKALPAKF 720

QY 3453 TEGLRNEEAVEGATAMLCELSKVAPVWRKGPENLRDGDYRCLRDGAVCELOICGLAM 3512
 Db 721 TEGLRNEEAVEGATAMLCELSKVAPVWRKGPENLRDGDYRCLRDGAVCELOICGLAM 780

QY 3513 ADAGEYLCVCGQERTSATLITRALPARFIEDVKNQEARREGATAVLQCELNSAAPVWRKG 3572
 Db 781 ADAGEYLCVCGQERTSATLITRALPARFIEDVKNQEARREGATAVLQCELNSAAPVWRKG 840

QY 3573 SETLRDGDYRSLRDGKTCELQIRGLAMADTGEYSCVCGQERTSAMLTVRALPIKFTGEL 3632
 Db 841 SETLRDGDYRSLRDGKTCELQIRGLAMADTGEYSCVCGQERTSAMLTVRALPIKFTGEL 900

QY 3633 RNEEATEGATAVLRCELSKMAVPEVWKGHETLRDGDHSLRDGAVCELOICGLVAEDAG 3692
 Db 901 RNEEATEGATAVLRCELSKMAVPEVWKGHETLRDGDHSLRDGAVCELOICGLVAEDAG 960

QY 3693 EYLCMCGKERTSAMLTVRAMPKFTIEGLRNEEATEGDTATLWCELSKAAPVWRKGHETL 3752
 Db 961 EYLCMCGKERTSAMLTVRAMPKFTIEGLRNEEATEGDTATLWCELSKAAPVWRKGHETL 1020

QY 3753 RDGDHSLRDGSRCELOIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFIEDVKNQ 3812
 Db 1021 RDGDHSLRDGSRCELOIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFIEDVKNQ 1080

QY 3813 AREGATAVLQCELSKAAPVWRKGSSETLRDGDYRSLRDGKTCELQIHGLSVADTGEYSC 3872
 Db 1081 AREGATAVLQCELSKAAPVWRKGSSETLRDGDYRSLRDGKTCELQIHGLSVADTGEYSC 1140

QY 3873 VCGQERTSATLTVR----- 3886
 Db 1141 VCGQERTSATLTVRALPARFTQDLTKAESEGATATLQCELSKAPVWRKGPETLRDGG 1200

QY 3887 ----- 3886

Db 1201 RYSLKQDGTFCLELOIHDLVADAGEYSCVCGQERTSATLTVRALPARFTEGLRNEEAMEG 1260

QY 3887 ----- 3886

Db 1261 ATATLQCELSKAAPVWRKGSSETLRDGDYRSLRDGAVCELOIHGLAMADNGVYSCVCGQ 1320

QY 3887 ----- 3886

Db 1321 ERTSATLTVRALPARFIEDMRNOKATEGATVTLQCKLRKAAAPVWRKGPNTLKDGRYSL 1380

QY 3887 ----- 3886

Db 1381 KQDGTSCLEQIRGLVIADAGEYSCICEQERTSATLTVRALPARFIEDVRNHEATEGATAV 1440

QY 3887 ----- 3886

Db 1441 LQCELSKAAPVWRKGSSETLRDGDYRSLRDGKTCLEQIRGLAVEDTGEYLCVCGQERTS 1500

| FT | NON TER | 1596 | 1596 |
|---------|-----------------------|--|------|
| SQ | SEQUENCE | 1596 AA; 89937E0760C4E2BF CRC64; | |
| | Query Match | Score 6619.5; DB 4; Length 1596; | |
| | Best Local Similarity | 81.7%; Pred. No. 0; | |
| | Mismatches | Conservative 70; Mismatches 203; Indels 21; Gaps 7; | |
| QY | 2726 | RARVRHDLHVGIITKRLKTMEVLEGESCSPECVLSHESASDPANWTVGGTKVGSSSRFOA | 2785 |
| DB | 1 | RARVRHDLHVGIITKRLKTMEVLEGESCSPECVLSHESASDPANWTVGGTKVGSSSRFOA | 60 |
| QY | 2786 | TROGRKYILVVREAAAPSDAGVEVFVRGLTSKASLI VRERPAALIKPLEDQWVAPGDVE | 2845 |
| DB | 61 | TROGRKYILVVREAAAPSDAGVEVFVRGLTSKASLI VRERPAALIKPLEDQWVAPGDVE | 120 |
| QY | 2846 | LRCELSRAGTPVHMLKORKAIRKSQKYDVVCEGTMAMLVIRGASLKADGETCEVERASKS | 2905 |
| DB | 121 | LRCELSRAGTPVHMLKORKAIRKSQKYDVVCEGTMAMLVIRGASLKADGETCEVERASKS | 180 |
| QY | 2906 | TASLHVEKANCFTTELTNLQVEEKGTAVFTCKTEHPAAVTWRKGLELRASGHQPQS | 2965 |
| DB | 181 | TASLHVEKANCFTTELTNLQVEEKGTAVFTCKTEHPAAVTWRKGLELRASGHQPQS | 240 |
| QY | 2966 | EGLTLRLTISALEKADSTYTCDIGQAQRAQLLVQGRRVHIIEEDLDVDVOEGSSATFR | 3025 |
| DB | 241 | EGLTLRLTISALEKADSTYTCDIGQAQRAQLLVQGRRVHIIEEDLDVDVOEGSSATFR | 300 |
| QY | 3026 | CRISFANYEPVHWFLDKTPLHANELNEIDAOPGVHVLTQLQALKDSGITTFEAGDORA | 3085 |
| DB | 301 | CRISFANYEPVHWFLDKTPLHANELNEIDAOPGVHVLTQLQALKDSGITTFEAGDORA | 360 |
| QY | 3086 | SAALRVTEKPSVFSRELTDATITEGEDTLVCETSTDCIPWCWTOKGTLRGSRQCULSH | 3145 |
| DB | 361 | SAALRVTEKPSVFSRELTDATITEGEDTLVCETSTDCIPWCWTOKGTLRGSRQCULSH | 420 |
| QY | 3146 | EGHRAQLLITGATLODSGRYKCEAGGACSSIVRVHARPVPFFQALKDLEVLEGGAATLR | 3205 |
| DB | 421 | EGHRAQLLITGATLODSGRYKCEAGGACSSIVRVHARPVPFFQALKDLEVLEGGAATLR | 480 |
| QY | 3206 | CYLSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVRNLRPODSGRYSCSGFDQTSTA | 3255 |
| DB | 481 | CYLSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVRNLRPODSGRYSCSGFDQTSTA | 540 |
| QY | 3266 | TUTVTALPAQFIGKLNRKEATGATATLRCELSTAPVEMKGSSETLRDGRYCLRDQGA | 3325 |
| DB | 541 | TUTVTALPAQFIGKLNRKEATGATATLRCELSTAPVEMKGSSETLRDGRYCLRDQGA | 600 |
| QY | 3326 | MCELOIRGLAMVDAAEYSCVCGEERTSASTIRPMPAHFIGLRHQESI EGATATLRCEL | 3385 |
| DB | 601 | MCELOIRGLAMVDAAEYSCVCGEERTSASTIRPMPAHFIGLRHQESI EGATATLRCEL | 660 |
| QY | 3386 | SKAAPVEMKGBESLRDGRHSRLRQDGAVCELOICGLAVADAGEYSCVCGEERTSATITV | 3445 |
| DB | 661 | SKAAPVEMKGBESLRDGRHSRLRQDGAVCELOICGLAVADAGEYSCVCGEERTSATITV | 720 |
| QY | 3446 | KALPAKFTGLELRNEBAVEGATAMLWCELSKVAPVEMKGPENLRDGRYILRQEGRCEL | 3505 |
| DB | 721 | KALPAKFTGLELRNEBAVEGATAMLWCELSKVAPVEMKGPENLRDGRYILRQEGRCEL | 780 |
| QY | 3506 | QICGLAVADAGEYLCVCGOERTSATITRALPARFIEDVKQEBAREGATAVLOCELNSAA | 3566 |
| DB | 781 | QICGLAVADAGEYLCVCGOERTSATITRALPARFIEDVKQEBAREGATAVLOCELNSAA | 840 |
| QY | 3566 | PVEWRKGSSETLRDGRYSRLRQDGTKCELQIRGLAMADTGEYSCVCGEERTSAMLTRALP | 3625 |
| DB | 841 | PVEWRKGSSETLRDGRYSRLRQDGTKCELQIRGLAMADTGEYSCVCGEERTSAMLTRALP | 900 |
| QY | 3626 | IKFTEGLRNEEATEGATAYLRCELSCMAPVEMKGMHETLRDGRHSRLRQDGARCBLQIRG | 3685 |
| DB | 901 | IKFTEGLRNEEATEGATAYLRCELSCMAPVEMKGMHETLRDGRHSRLRQDGARCBLQIRG | 960 |
| QY | 3686 | IYAEADAGEYLCMGKERTSAMLTRAMPBSKETEGRLRNEATEGDTATWLCELSKAAPVEM | 3745 |
| DB | | | |
| NON TER | | | |

Db 961 LVAEDAGEYLWCGKERTSAMLTVRAMPKSFIEGLURNEATEGDTATLWCBLSKAAAPVW 1020
 QY 3746 RKGHETLRDGRHSRQDGRSCELRQIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFI 3805
 Db 1021 RKGHETLRDGRHSRQDGRSCELRQIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFI 1080
 QY 3806 EDVKQEARBEGATVLOCELSKAAAPVWVRKSGSETLRGDRYSLRGDGRTRCELRQIRGLHGSVA 3865
 Db 1081 EDVKQEARBEGATVLOCELSKAAAPVWVRKSGSETLRGDRYSLRGDGRTRCELRQIRGLHGSVA 1140
 QY 3866 DTGEYSCVCGQERTSATLTVRALPARFIQVREPELQSLQAEBSGATLQCELSPTATVWVSKG 3925
 Db 1141 DTGEYSCVCGQERTSATLTVRALPARFIQVREPELQSLQAEBSGATLQCELSK-VAPVWRKG 1199
 QY 3926 GLQIANGREPRLOGCTAELVLODQREDTGEYTCGSGQATSATLTVTAAPVRFLEL 3985
 Db 1200 PETLRDGGYSLKQDGRTRCELRQIRGLHDSVADAGEYSCVCGQERTSATLTVRALPARFTEGL 1259
 QY 3986 OHQEVDEGTAHLCCELSPAGASVWVRKSGSLQFPCKAYQWQDGAALVVRGVEDA 4045
 Db 1260 RNEEAMEGATVLOCELSKAAAPVWVRKSGLEALRDGDKYSLRGDGRTRCELRQIRGLHGSVA 1318
 QY 4046 GDYTCDTGHTOSMASLSVRPRPKTRLOSLQEQETGDIARLCCQLSDASSGAVVOMLKE 4105
 Db 1319 GYSCVCGQERTSATLTVRALPARFIQVREPELQSLQAEBSGATLQCELSKAAAPVWVRKSG 1375
 QY 4106 GVHLAGPKYMRSGATRELLIHQLEAKDTGEYACVTCGQKTAASLRVTEPEVTIVRGL 4165
 Db 1376 PNTLKGDRYSLKQDGRTRCELRQIRGLHDSVADAGEYSCVCGQERTSATLTVRALPARFIEDV 1435
 QY 4166 VDAEYTADEDFEFCSEVRAGATGQWCLQPLQSNTEVEVAVRDGRIHTLRKGVTP 4225
 Db 1436 RNHEATEGATVLOCELSKAAAPVWVRKSGSETLRDGRYSLR-QDTRCELRQIRGLAVE 1492
 QY 4226 DAGTVSFLGNHSAQITVRAPEVTILEPLQDVLQSEGQDASQCKLSRASQEARWAL 4285
 Db 1493 DTGEYLCVCGQERTSATLTVRALPARFIQVREPELQSLQAEBSGATLQCELSKAAAPVWVRKSG 1550
 QY 4286 GGVLQANEMNDITVEQGTLLHLLTKHVTELDAGTVSFHVGC 4328
 Db 1551 GPEILRDGRHSRQDGRSCELRQIRGLHDSVADAGEYSCVCGQERTSATLTVRALPARFIEDV 1592

RESULT 5

Q8NH2 PRELIMINARY; PRT; 1040 AA.
 ID Q8NH2
 AC Q8NH2;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Obscurin (fragment).
 GN OBSN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Young P., Ehler E., Gautel M.;
 RT "Obscurin, a giant sarcomeric Rho-GFP protein involved in sarcomere assembly";
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ314506; CAC85753.1;
 DR InterPro; IPR003962; FNIII_subd.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR003599; IG.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003598; IG_c2.
 DR InterPro; IPR000048; IG-region.
 DR Pfam; PF00041; fn3; 1.
 DR Pfam; PF00047; ig; 6.
 DR Pfam; PF00612; IQ; 1.

DR PRINTS; PR00014; FNTYPEIII.
 DR SMART; SM00060; FN3; 1.
 DR SMART; SM00409; IG; 6.
 DR SMART; SM00408; IG2; 5.
 DR SMART; SM00015; IQ; 1.
 DR PROSITE; PS00835; IG_LIKE; 5.
 DR PROSITE; PS0096; IQ; 1.
 KW Immunoglobulin domain; Repeat.
 FT NON_TER 1
 FT TER 1040 1040
 SQ SEQUENCE 1040 AA; 112354 MW; 3DE879CA266D7F44 CRC64;
 Query Match 13.0%; Score 5358.5; DB 4; Length 1040;
 Best Local Similarity 99.4%; Pred. No. 2.6e-279;
 Matches 1034; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
 QY 4430 WRLEILEPLKNAAVRAGAAQARTCTLSEAVPVGEASWYINGAAVQPDSDSDWTTADGSHQ 4489
 Db 1 WRLEILEPLKNAAVRAGAAQARTCTLSEAVPVGEASWYINGAAVQPDSDSDWTTADGSHH 60
 QY 4490 ALLLSAQPHHAGEVTFACRDVAVASARLTVLGLPDPPEDAEVAHSSHTVLSWAAPMSD 4549
 Db 61 ALLLSAQPHHAGEVTFACRDVAVASARLTVLGLPDPPEDAEVAHSSHTVLSWAAPMSD 120
 QY 4550 GGGGLCGYRVEKGAQGWRLCHLVGPEPCVDGLAPGETYRPRVAAGVPVGAAGEPVH 4609
 Db 121 GGGGLCGYRVEKGAQGWRLCHLVGPEPCVDGLAPGETYRPRVAAGVPVGAAGEPVH 180
 QY 4610 LPQTVRLAEPKVPVQPSAPESROVAAGEDVSLVLEVAAGVIVHKGHERIQPGRF 4669
 Db 181 LPQTVRLAEPKVPVQPSAPESROVAAGEDVSLVLEVAAGVIVHKGHERIQPGRF 240
 QY 4670 FVVSQGRQMLVIKFTADQGEYHGLAQSICPAATFOVALSPASVDPAQPSLPPE 4729
 Db 241 FVVSQGRQMLVIKFTADQGEYHGLAQSICPAATFOVALSPASVDPAQPSLPPE 300
 QY 4730 AAQEGDLHLLWEALARKMRSEPTLDSISLPEEDGRSQRLOFAEVEADPLSEGYSTA 4789
 Db 301 AAQEGDLHLLWEALARKMRSEPTLDSISLPEEDGRSQRLOFAEVEADPLSEGYSTA 360
 QY 4790 DELARTGDADLSHTSSDDESRAQTSLVLYLKAGRPCTSPASKVGAAPSVKPPQQQ 4849
 Db 361 DELARTGDADLSHTSSDDESRAQTSLVLYLKAGRPCTSPASKVGAAPSVKPPQQQ 420
 QY 4850 EPLAAVRPPLGDLSTKDLGDPMDKAAVKIQAAKGYKVRKEMKQOQSGPFSHTFGTEA 4909
 Db 421 EPLAAVRPPLGDLSTKDLGDPMDKAAVKIQAAKGYKVRKEMKQOQSGPFSHTFGTEA 480
 QY 4910 QVGDALRLCEVAVSKADVRAEWLKDGVVELTDGRHHHIDQLGDGTCSLIIAGLDRADAGCY 4969
 Db 481 QVGDALRLCEVAVSKADVRAEWLKDGVVELTDGRHHHIDQLGDGTCSLIIAGLDRADAGCY 540
 QY 4970 TCQVSNKFGQVTHSACVVSSESESSSGELDDAFRAARLHLRLFRTKSPAESVDE 5029
 Db 541 TCQVSNKFGQVTHSACVVSSESESSSGELDDAFRAARLHLRLFRTKSPAESVDE 600
 QY 5030 ELFLSADGPAEPPEADWQTYREDEHFCIRFALTEARQAVTRFOEMFATLIGIVEIK 5089
 Db 601 ELFLSADGPAEPPEADWQTYREDEHFCIRFALTEARQAVTRFOEMFATLIGIVEIK 660
 QY 5090 LVEQGRPRVEMCISKETPAPVVPPEPLPSLLTSDAAPVFLTELQNVQGVQGVPSFPCVV 5149
 Db 661 LVEQGRPRVEMCISKETPAPVVPPEPLPSLLTSDAAPVFLTELQNVQGVQGVPSFPCVV 720
 QY 5150 TGQPMPSVRWPKDGLLEDDHYMINEDQGGHOLIITAVVPADMVYVRCIAENSMGVSS 5209
 Db 721 TGQPMPSVRWPKDGLLEDDHYMINEDQGGHOLIITAVVPADMVYVRCIAENSMGVSS 780
 QY 5210 TKAEVLRLTSTDYDTAADATESSSYSAQGYLS-SREQEGTSTTDEGQLPQVVEELRD 5268
 Db 781 TKAEVLRLTSTDYDTAADATESSSYSAQGYLS-SREQEGTSTTDEGQLPQVVEELRD 840
 QY 5269 LQVAPGTRLAKFQLKVKGYPAAPRLYWFKQOPLTASAHRTMTCKKILHTLEIISVTRDS 5328

| | | | |
|----------|-------------|---|---|
| Db | 841 | LQVAPGRLAKFQLKVKGYPAPDYLPFKOCQPLTASAHIRMTDKKILHTLEIISVTREDS | 900 |
| Qy | 5329 | GOYAAYISNANGAAYSSARLLVVRGPDPPEKPSADVHEQLVPPRMLERFTPKVKXGSSI | 5388 |
| Db | 901 | GOYAAYISNANGAAYSSARLLVVRGPDPPEKPSADVHEQLVPPRMLERFTPKVKXGSSI | 960 |
| Qy | 5389 | TFSVKVEGRPVPTVHMLREAEERGLVIGDPDTFGYTVASSAQOHSVLVLDVGSQHOGTYT | 5448 |
| Db | 961 | TFSVKVEGRPVPTVHMLREAEERGLVIGDPDTFGYTVASSAQOHSVLVLDVGSQHOGTYT | 1020 |
| Qy | 5449 | CIASNAAGQALCSASLHVSG | 5468 |
| Db | 1021 | CIASNAAGQALCSASLHVSG | 1040 |
| RESULT 6 | | | |
| Q86YC7 | | PRELIMINARY; | |
| ID | Q86YC7 | PRELIMINARY; | PR7 767 AA. |
| AC | Q86YC7 | | |
| DT | 01-JUN-2003 | (TREMBlrel. 24, Created) | |
| DT | 01-JUN-2003 | (TREMBlrel. 24, Last sequence update) | |
| DT | 01-OCT-2003 | (TREMBlrel. 25, Last annotation update) | |
| DE | | Similar to titin (Fragment). | |
| OS | | Homo sapiens (Human). | |
| OC | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC | | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| OX | | NCBI_TaxID=9606; | |
| RN | | [1] | |
| RP | | SEQUENCE FROM N.A. | |
| RC | | TISSUE=Placenta; | |
| RA | | Strausberg R.; | |
| RL | | Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases. | |
| DR | | EMBL; BC043916; AAH43916.1; -. | |
| DR | | InterPro; IPR003599; IG. | |
| DR | | InterPro; IPR007110;-IG-like. | |
| DR | | InterPro; IPR003598; IG_c2. | |
| DR | | Pfam; PF00047; ig; 5 | |
| DR | | SMART; SMC0409; IG; 7. | |
| DR | | SMART; SMC0408; IGc2; 5. | |
| DR | | PROSITE; PS50835; IG_LIKE; 5. | |
| FT | | NON TER | 1 |
| SQ | | SEQUENCE | 767 AA; 84732 MW; 39AB316B63964DAA CRC64; |

| | | | | |
|-----------------------|-----------------|--|-----------|-------------|
| Query Match | 9.4%; | Score 3865; | DB 4; | Length 767; |
| Best Local Similarity | 99.5%; | Pred. No. 3.7e-199; | | |
| Matches 751; | Conservative 3; | Mismatches 1; | Indels 0; | Gaps 0; |
| Qy | 2248 | VKVQGRVYTLIRVLAEDAGEIQFVAENAESRAQLRVKELPVTILRPLRDKTAMEXKRG | 2307 | |
| Db | 5 | IRQGRYTLIRRVLAEDAGEIQFVAENAESRAQLRVKELPVTILRPLRDKTAMEXKRG | 64 | |
| Qy | 2308 | VLEQVSRASAQVRFKGSQBLQCPKYELVSDGLYRKLIIISDVHAEDEDTYTCDAGNVK | 2367 | |
| Db | 65 | VLEQVSRASAQVRFKGSQELQCPKYELVSDGLYRKLIIISDVHAEDEDTYTCDAGNVK | 124 | |
| Qy | 2368 | TSAQFFVEEQSIITIVRGLODVTVMEPAPAFWPECETSI PSVRPPKWLIGKTVLQAGNVGL | 2427 | |
| Db | 125 | TSAQFFVEEQSIITIVRGLODVTVMEPAPAFWPECETSI PSVRPPKWLIGKTVLQAGNVGL | 184 | |
| Qy | 2428 | EQEGTVHRLMLRRTCSMTGSPVHTVTKGSRSSARLVVSDIPVILTRPLEPKTGRELSVV | 2487 | |
| Db | 185 | EQEGTVHRLMLRRTCSMTGSPVHTVTKGSRSSARLVVSDIPVILTRPLEPKTGRELSVV | 244 | |
| Qy | 2488 | LSCDFRPAPKAVQVYKDDTPI SPSEKFKQSLLEGQVAELIURLMPADAGVYRCAQSAHS | 2547 | |
| Db | 245 | LSCDFRPAPKAVQVYKDDTPI SPSEKFKQSLLEGQVAELIURLMPADAGVYRCAQSAHS | 304 | |
| Qy | 2548 | STEVTVAREVTVTGPLQDAATEEGWASFCSEL SHEDEEVESWINGMPLVNDSFHISH | 2607 | |
| Db | 305 | STEVTVAREVTVTGPLQDAATEEGWASFCSEL SHEDEEVESWINGMPLVNDSFHISH | 364 | |
| Qy | 2608 | KGRRHLLVKSIIQRADAGIVRASSLKIVSTARSARLEVRVPKVVFLLKALDDLSAEERGTLALQ | 2667 | |

| | | | |
|----------|---|--|------|
| Db | 365 | KGRHILVLKSIQRADAGIVRASLKVSTSRLEVAVKPVFLKALDDLSAERGTIALQ | 424 |
| Qy | 2668 | CEVSDPEAHVVRKQGVOLGSPDKYDPLHTAGTRGLVHVHVSPEADAGLYTCHVGSETRA | 2727 |
| Db | 425 | CEVSDPEAHVVRKQGVOLGSPDKYDPLHTAGTRGLVHVHVSPEADAGLYTCHVGSETRA | 484 |
| Qy | 2728 | RVRVHDIHVGTIKELKTMEVLEGESCFEVLHSHESADPMATVGGKTVGSSSRFOATR | 2787 |
| Db | 485 | RVRVHDIHVGTIKELKTMEVLEGESCFEVLHSHESADPMATVGGKTVGSSSRFOATR | 544 |
| Qy | 2788 | QGRKYIILVREAAPSDAGEVVFVSRGLTSKASLIVRERPAATIKPLEDQWVAGEDVELR | 2847 |
| Db | 545 | QGRKYIILVREAAPSDAGEVVFVSRGLTSKASLIVRERPAATIKPLEDQWVAGEDVELR | 604 |
| Qy | 2848 | CELSRAGTPVHMLKDRKAIRKSKYDVVCGTWMLVIRGASLKDAGEYTCVEASKSTA | 2907 |
| Db | 605 | CELSRAGTPVHMLKDRKAIRKSKYDVVCGTWMLVIRGASLKDAGEYTCVEASKSTA | 664 |
| Qy | 2908 | SLHVEEKANCFTEBLTNLQVEEKGTAFTVCTKTEHPAATVTRKGLLELRASGKHQPSQEG | 2967 |
| Db | 665 | SLHVEEKANCFTEBLTNLQVEEKGTAFTVCTKTEHPAATVTRKGLLELRASGKHQPSQEG | 724 |
| Qy | 2968 | LTLRLTISALEKADSDTYTCDIGQASRAQLLVQG | 3002 |
| Db | 725 | LTLRLTISALEKADSDTYTCDIGQASRAQLLVQG | 759 |
| RESULT 7 | | | |
| Q8WZB3 | PRELIMINARY; PRT: 26926 AA. | | |
| ID | Q8WZB3 | | |
| AC | Q8WZB3; | | |
| DT | 01-MAR-2002 (TrEMBLrel. 20, Created) | | |
| DT | 01-MAR-2002 (TrEMBLrel. 20, Last sequence update) | | |
| DT | 01-OCT-2003 (TrEMBLrel. 25, Last annotation update) | | |
| DE | N2B-titin isoform. | | |
| GN | TTN. | | |
| OS | Homo sapiens (Human). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Carnivora; Insectivora; Hominoidea; Homo. | | |
| OX | NCBI TaxID=9606; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | MEDLINE=20309627; PubMed=10850961; | | |
| RA | Freiburg A., Trombetta K., Hall W., Casoria O., Pougourousse F., | | |
| RA | Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C., | | |
| RA | Granzier H., Labeit S.; | | |
| RT | "Series of exon-skipping events in the elastic spring region of titin | | |
| RT | as the structural basis for myofibrillar elastic diversity."; | | |
| RL | Circ. Res. 86:1114-1121(2000). | | |
| RN | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=21573839; PubMed=11717165; | | |
| RA | Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M., | | |
| RA | McNabb M., Witt C.C., Labeit D., Gregorio C.C., Granzier H., | | |
| RA | Labeit S.; | | |
| RT | "The complete gene sequence of titin, expression of an unusual ~700 | | |
| RT | kDa titin isoform and its interaction with obscurin identify a novel | | |
| RT | Z-line to I-band linking system."; | | |
| RL | Circ. Res. 89:1065-1072(2001). | | |
| CC | -!- SIMILARITY: CONTAINS 91 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS. | | |
| DR | EMBL; AJ277892; CAD12455.1; - | | |
| DR | GO; GO:0016020; C:membrane; IEA. | | |
| DR | GO; GO:0005524; F:ATP binding; IEA. | | |
| DR | GO; GO:0004886; F:hematopoietin/interferon-class (D200-domain. .; IEA. | | |
| DR | GO; GO:0004601; F:peroxidase activity; IEA. | | |
| DR | GO; GO:0004674; F:protein serine/threonine kinase activity; IEA. | | |
| DR | GO; GO:0004713; F:protein-tyrosine kinase activity; IEA. | | |
| DR | GO; GO:0016740; F:transferase activity; IEA. | | |
| DR | GO; GO:0005975; P:carbohydrate metabolism; IEA. | | |
| DR | GO; GO:0006468; P:protein amino acid phosphorylation; IEA. | | |
| DR | GO; GO:0006999; P:response to oxidative stress; IEA. | | |
| DR | InterPro; IPR000282; Cytochrome_c_receptor_2 | | |

Db 2658 V---BAVKIKKTL-KNLTVTBT-QDAVFTVBLTHPNVKGWIKNGVVLNESNEKALISVK 2712
 Qy 1310 GCMRLVVQOAGADAGEYTCAG--GQRLSHLD---VSEPKAVFAKEQLAHRKVQAE 1363
 Db 2713 GTIYSLRINKCAIVDES VYGRGLGASARLHVETVKIIKKPKDVTALN----- 2763
 Qy 1364 AGATATLCEVAQAQ--TEVTWYKDGKLSKSSKVRMEAVGCTRLVVQOACQADTGEYS 1421
 Db 2764 ---ATVAFEVSVSHDTPVVKFHKNVKIKPSDKHRLVSEKVKHKLMLQNLSPSDAGEYT 2819
 Qy 1422 CEAGGQRLSFLDVAEPKVPFAKQPVH-----REVQAQAGASTTSLCEVAQAQTEVMWY 1476
 Db 2820 AVVGQLECKAKLFV-----ETLHITKTMKNIEVPTKTASFCEVSEVSHFVPSMWL 2869
 Qy 1477 KDGKLSKSSKVRMEAVGCTRLVVQOAGADAGEYSCAQSRLSHLHVFAEKAVFAK 1536
 Db 2870 KNGVEIEMSEKFI VVQKGLHQLIIMNTSTEDSAEYTFVCGNDQVSATLVT-PMITS- 2927
 Qy 1537 EOPASREVQAAGTATLSCEVAQAQTEVTWYKDGKLSKSSKVRMEAVGCTRLVVQEA 1596
 Db 2928 --MLKDIANAEEKDTITEVTWYEGISYKWLKNGVEIKSTDKQMRTKLTHSLNRN 2984
 Qy 1597 GOADAGEYSCAGDORLSFHLHVABPKVFAKEQPAHREVQAQAGASATLSCEVAQAQTE 1656
 Db 2985 HFGDAADYTFVAGKATSTATLYVEARHIEFRKH---INKDIKLEKGRAMFECEVSEPDIT 3041
 Qy 1657 VTWYKDGKLSKSSKVRMEAVGCTRLVVQOAGADAGEYSCAQSRLSHLHVFAELP 1716
 Db 3042 VQWMDGDELQITDRIKQKQYVHRLIIPSTRSDAGKITVAVAGNVSTAKL FVEGRDV 3101
 Qy 1717 QISERPCRRPLWKEHEDIIILATLATPSAATVWLKDGVEI--RRSKRHETASQGDTH 1774
 Db 3102 RI--RSIKKEVQVIEKQAVV--BEVNEDDVAHWYKDGIEINFQVERHKYVVERRIH 3157
 Qy 1775 TLTVHGAQVLDSATYSCVRAEGQDPVQVBE-EVAAKFCRLLEBVCGLGCTVTLACELS 1833
 Db 3158 RMFISETRSDAGEYTFVAGNRSSVTLVYNAPEPPQVLQBPVTVQSG-----K 3208
 Qy 1834 PA--CAEVVMRCGNTQPRV-----GKRFQVMAEGPVRSLTVLGLRAEDAGBYVC 1880
 Db 3209 PARFCAVI---SGRPQPKISWYKEBQLSTGFKCKFLHDGQGYTLILLIEARPEDAAYTC 3265
 Qy 1881 ESRDHP-----TSAQITVSVPRVVK-----FMGLSTVVAEBEGEATFCQCVSPS 1925
 Db 3266 BAKNDYGVATTSSASLSVPEVSPQEMPVYPAITFLQDVTVTSEGPAPFCQVSGT 3325
 Qy 1926 DVAVVWFRDQALQSEKFAISQSGASHLATISDLVLEDAGQITVEABGA---SSSAAL 1981
 Db 3326 DLKVSWSKDKKISRFRTQFEDTYQLBIAEAYPEDEGTYTFVASNAGVQVSSANL 3385
 Qy 1982 RVREAPVLFKKLEPQ-----TVEERSSTVLEVELTR----- 2013
 Db 3386 SLEAPESILHERIEQIEEMEMKEFSSFLSAEEGLHSAELQLSKINETLELLGESPVYS 3445
 Qy 2014 -----PWELRWTRNATALAPKKNVEI 2035
 Db 3446 TKFDESEKGTGPIFTKEVSNADISMGDVATLSVTVIGIPKPKIOFFNGVLLTSPADYK 3505
 Qy 2036 HAEGARHRLVLNVGFAIRGFGCETPD----- 2066
 Db 3506 VFDGDDHSLIILFTKLEDEGEYTCNASNDYKTCISAYLKINSKGEHGKDTETESAVAKS 3565
 Qy 2067 QAKL-----TVEMRQVRLVRLGQAV-EAREQGTATWEVOLSHADVDGGSWTDRGLRFQ 2117
 Db 3566 LEKLGCPCPPHFLKELKPIRCAQGLPAIFETVVGEPAPTV-----TWFKENKOL- 3615
 Qy 2118 QGPTC-----HLAVRGP--MHTLTLGLRPDSGLMVVFAEGV---HTSARLVV----- 2161
 Db 3616 ----CTSYYTTHPNPNSGTFVNDPQREDSGLYICAKENMLGESTCAEALLVLEDDTD 3671
 Qy 2162 -----TELPVFSR-PL-----QDVVTTEKHKVTLCELSRPNVDVRW 2198
 Db 3672 MTDTPCAKSTPEAEDFPQELPKGPAVEALDSEQEIATFVKDTILKAALITEENQOLSY 3731

Qy 2199 -----LKOGVELPAGKTMATAAOG-----ACRSITY----- 2225
 Db 3732 EHIKANELSSQLPLGAQELQILEQDKLTPESTREFLCINGSIHFPQLKEPSPNLQLOI 3791
 Qy 2226 ---RCFADQGVVVCDAHDAQS----- 2244
 Db 3792 VQSQKTFSEKGLMPPEPETQAVLSTEKI FPSAMSI EQINSITVEPLKTLAEPGNYP 3851
 Qy 2245 SASVKVQGRTY-----TLIVRRVLAE----- 2265
 Db 3852 QSSIEPPMHSYLTSAEVLSPKEKTVSTNREQRTVLOKQBAQSALILQSILAEHVES 3911
 Qy 2266 -----DAGEIQFVA-----ENAESRAQLRVKE----- 2287
 Db 3912 LQSPDVMISQVNVPELVPSEHSCTEGKILISANPLENAGQDSAVRIEGLKSLRPLAL 3971
 Qy 2288 -----LPVTLVRPLRDXIAMKRGV----- 2308
 Db 3972 EEKQVLKEEHSNVMVPPDQIIESKREPVAIKKQVEQVQGRDLLSKESLLSGIPEQRNL 4031
 Qy 2309 LECQVSR-ASQV-----RW----- 2322
 Db 4032 LKIQICRALQAAVASQPGLFSEWLRNIEKVEAVNIITQEPRHIMCMYLVTSKASVTSE 4091
 Qy 2323 -----FKGSO 2327
 Db 4092 VTIIIEDVDPQANLXWELRDALCAIYEEIDILTAEGPRIQOGAKTSLQEEMDSFGSQ 4151
 Qy 2328 ELOGPKYELVSDGL-----YKLIISDVH-----AE 2354
 Db 4152 KVPEITEPEVESKYLISPEEVSFNVQSVKYLDATPVTGASAVVSDEKQDESUKPSE 4211
 Qy 2355 DEBTVTCDAQ-----DVKTSQAFVEBOSITIVRGLQDVTWMEPAPAFECETSI PSV 2407
 Db 4212 EKESSESSEGTETAVATVKEABGGIKEDGPMIHTPLVD-TVSEEGDI-VHLLTITNA 4269
 Qy 2408 RPPKWLGLGTVLOAGNVGLQEGTVHRLMLRR-----TCSTMTGPVHVTGKSR 2457
 Db 4270 KEVWYFENKLPSPDEKFKCLQDQNTYTLVIDKVNTEHQGEVCEALNDS-----GKTA 4324
 Qy 2458 SSARL-VVSDI PVLTPLPTKGTRELQSVLSCDPRPAPKA-VQWYKDDTPLSPSEK 2515
 Db 4325 TSAKLTWVRAAPVIRKIEPLEVALGHLAKTCEIQSAPNVRFQFKAGRIEYSDKGS 4384
 Qy 2516 MSLEGQMAELIIRLMPADAVYRCQAGSAHSESTVEAREVTVTG-----PLQD 2566
 Db 4385 IRSKYSIISLEILRTQVDCGEYTCASNEYSVSCTA-TLTVTVPGGEKKYKLLPERK 4443
 Qy 2567 ABATEGWNASFCELSHEDBEVWSLNGMPLNDSFHEISHKGRHTLVILKIQRADAGI 2626
 Db 4444 PEPKEE--VVLKSVLSEKRPSEEPKVEPKL-----EKVKPAPVPEPPPKPVEVEVPT 4496
 Qy 2627 VRASSLKVSTARSLEVRVVPVFLKALDDLSABERGTLALQCEVSDPE-----AHV 2677
 Db 4497 VTKREKIPEPTKVP-EIKPAIPLPAPKPKPEAEVTKIKPPVEPETPIAPVTVFV 4555
 Qy 2678 VWRKGVQGLSPDKYDFLHTAG-TRGLVVDVSPEDAGLYTCHVGS-----BETRARVR 2730
 Db 4556 VGKKAQAK--APKEBAAPKGPBKIGVPKKTSPIDAEERKLRPGSGGKPPDEAPFTYQ 4612
 Qy 2731 VHDLVHGTIKRLKTMVLE-----GESCSPECVLSHESASDPAMWTGGTGVSSSSFOAT 2786
 Db 4613 LKAVPLKFKVKEIKDIITLSEFVGSSEIPECLVSPSTAI--TTWMDGKNIESPKHRFI 4670
 Qy 2787 ROGRKYLIVREAPSDAGEVWFSVR-----GLTSKASLIVRERPAIILKPLEDQ-WVAPG 2841
 Db 4671 ADGKDRKLHIIDVQLSDAGEYTCVLRGNKEKTSTAKLVEELVFRVFKTLEBEEVWVG 4730
 Qy 2842 EDVELCELSRAGTTPHVLKORK-ATRKQKYDVVCEGTWAMLVIGASIKDAGEYTCV 2900
 Db 4731 QPLYLSCENLKE-RDVVWRKDGKIVVEKPGRIVPGVIGLMRALTINDADTDAGTIVTV 4789

Db 6884 -----LGPLTADDAFVEPTMDLSAFKDGLEIVVNPITILVPSTGY 6924
Qy 4887 KVRKEMKQOEGPMRSHFTGDTAOGVGDALRELCVASKADVARWLKDGVELTIDGRHHI 4946
Db 6925 -----PRPTATWCFGDKVLETDGRVKMKTLSA----- 6951
Qy 4947 DQDGDGTSLLIAGLRADAGCYTCQVSN-----KFGQVTHS 4983
Db 6952 -----YAEVLVSPERSDKGIYTLKLENRVKTTISGEIDVNVVIAPSPAKELKFGDITKD 7005
Qy 4984 ACVVVSGSESAESSGGELEDODAFRAARRLHRFRFK----- 5021
Db 7006 SVHLT-----WEPDDDDGSPLTGYVVEKREVSRTWTWKVMDFTDLEFTVFDLVQKREYL 7061
Qy 5022 -----SPAEVDELEFLSADGPAEPEPADWQYRDEHFI----- 5058
Db 7062 FKVCARNKCGGEFAYV-DEPVNMSPTATVPDPENVKWRDTANSIFLTDWPPKNDGGS 7120
Qy 5059 -----CIR-----FEALTEARQAVTRFOE--MFATLIGIGVEIKLVEOGPRRVE 5099
Db 7121 RIKGYIVERCPRGSKWACGEPVAETKQVETGLEEGKWA-----YRVK 7165
Qy 5100 MCISKETAPVVPPEPLSLTSDAAPVFLTE--LQOQEVODGYVSPFDCVVTGQPMPSV 5157
Db 7166 ALNRQAGSKPRPTEIOQAVDTQEAPEIFLDVKLAGUTVRAGTKIELPATVTKPEPKI 7225
Qy 5158 RWFKDGKLEEDDHYMINEDQGGHQLIITAVPADMGVYRCLAEANGSVSSTKAELRVD 5217
Db 7226 TWTKADMILKQDKRITI-ENVPKSTVTIVDSKRSDTGTGTIIEAVNCGRATAVVEVNV- 7283
Qy 5218 LTSTDYDTAATSSSVFSAGYLSSREQGTSTTDEGQLQPVVEELRLQVAPGTRL 5277
Db 7284 LDKFGPPAFTDVTNESCILLTNWPPRDDGGS----- 7316
Qy 5278 AKFOLKVKGYPAPRLYKFKDQOPLTASAHIMTKGKILHTLEIISVTRDSGQVAAVSN 5337
Db 7317 -----KINYYVER-----RATDSEVWHKL--SSIVKDNFNKATKLIP 7352
Qy 5338 AWGAAYSSARLLVRGDFPEKPSADVHVEQLVPRMLERFPPKKVKGGS-SITESVKVE- 5395
Db 7353 NKEYIFRVAANMVGVEFVQASPIKAKYQFDPGPPTRLEPSSDITKDAVTLTWCEPDD 7412
Qy 5396 -GRPVPVTHLREAEERGVLMIGDPTPGYTVASSAQHSLVLLDVGRHQGTY--TCIAS 5452
Db 7413 GGSPI-TGYW-----ER-----LDPTDKVRCKNMPVKDTRYRVKGLTNKKYFRVLA 7463
Qy 5453 NAAGQALCASLHVSGLP-KYEEQEKVKEALISTFLOG--TQOISAGQLETSFADLGG 5509
Db 7464 NLAG-----PGKPSKSTEPILLKDPIDPPWPGKPTVKDVGKTSVA-----LNW 7507
Qy 5510 ORKEEPLAAKALHSLAEVGTBEFLQKLTISOITEMVSAKITQAKLQVPG-GPSDSDSK 5568
Db 7508 TKPEHDGAKIESYVIEMLKTGTDEWVR-----VAEGVPTTOHLLPGLMEGQBYSF 7558
Qy 5569 TPSAPRGRGRSPSSICSSSESDGDARGEIPIYVVTADYPLPLGAEOQDAITLREGQY 5628
Db 7559 RVRAVNKAGESEPSE-----PSDPVLCKREKLY 7595
Qy 5629 -----VEVLDAH---PLRWLVTKP-----TKSSPRQGH--VSPAYLDRRL 5666
Db 7596 PPSPPRLVINITKNTADLKWTPPEKDGSPITNYIIVEKRDVRRKGQVTDVTTVKOTKC 7645
Qy 5667 KLSR-ENG-----AAEAPFPGGAUSEDEYKARLSSVIOELLSSEQAFVEELQFLQSH 5718
Db 7646 TVTPTLGLSVYFRVAE-----NAIGQSDI-----TEIEDSVLAKDTFT----- 7695
Qy 5719 HLQHLERCPHPVIAVAGOKAVIFRNVDRIGRHFSSFLQELQOCDTDDVAMCFI-KNOAA 5777
Db 7696 -----TPGPPYALA-----VVDVTKRHVDLKWPEPKNDGGRPIQRYVIEKKERL 7729
Qy 5778 FEQYLEFLVGRV-----QAESVTVSTAIQEFYKYAE-----EALLAGDPS 5818
Db 7730 GTRWTK--AGKTAGPDCNFRVTDVIEGTEVO--FOVRAENEAGVGHSEPTEILSIEDPT 7795

Qy 5819 QPPPPPLQHYLE-----QPVRE-----VORYQALLKELIRNK 5850
Db 7786 SPSPPLDLUHVTDAGRKHIAIAWKPEKGGSPITGYHVEMCPVGTGRWNRVNSRPIKOL 7845
Qy 5851 ARNRQCALLEGAYAVVSALPORAENKLHVS--LMENYPG-----TLFALGEPIRQ 5899
Db 7846 KPKVEGVVVDKEY-----VLRVRAVNAIGVSPSEISENVAKDPCKCTID--LET 7896
Qy 5900 GHPIVWEG-----APGARMPWKGNRHV-----FLFRN-----HLVICKP-RED 5937
Db 7897 HDIIVTEGSKLSIPVFPRAVPVPTVSWHKDGKEVKASDLRTWKNDHISAHLEVPKSVRAD 7956
Qy 5938 SRDITVSYVFRNMKLSISDLNDQVEG-----DDRAFEV-----WQ----- 5973
Db 7957 AGIYITIT--LENKLGSAIASINVKVLGPGCKDIKASDITKSSCKLTWPEPFDDGGTPI 8014
Qy 5974 -----BRESVRKYLLOARTAIKSNW-VKETC-----GIQORLALPV 6010
Db 8015 LHVYLERREAGRRTYIPVMSGENKLSWTVDKILNPGEYFRVKAVNKVGGEGYIELKPV 8074
Qy 6011 -----WRPDP--PEELADCTAELGETVKLACRVGTQPKPVISWYKDGKAVQVDPHHILI 6063
Db 8075 IAQDPKQPPDPVDVEVHNPTAE-----AMTITWKPPL--YDGGSKIM--CYIIE 8120
Qy 6064 EDPDGS-----CALILSLTGVDSG--QYMCFAASAAGNCSITLKGILVQVPP-R 6109
Db 8121 KIAKGBERWKRCKNEHLVPIITYTAKGLEEGKEYQFVRAENAAG--ISEFSRATPTTK 8176
Qy 6110 FVNVKVASPVV-----EGEDAQFCTTIEGAPYQIRWYKD-----GALLT 6149
Db 8177 AVDPIDAPKVLRTSLEVVRGDEIADASISGSPYITITWIDENVIVBEEKRAAPLV 8236
Qy 6150 TGNKFO-----TLSEPRSGLLVLVIRAAASKEDGLYCELVNRLGSRAS 6194
Db 8237 RREKGEVQSEEPFVPLTQRLSIDNSKKGESQLRVSLSRDPHGLWIKVENDHGIKAP 8296
Qy 6195 AELRIOS-----PMLQAOEQCHREOLVAVE-----DTTLERAOE-----V 6231
Db 8297 CTVSVLDTGPPFINVFEDIRKTSVLCKWEPPLDDGSGSEIINYLEKKDKTKPDSEWIV 8356
Qy 6232 TSVLK-----RLLGPK-----APGPS-----TGDLTGPGPCP----- 6258
Db 8357 TSTLRCKSVTKLISGKEYLFRVRAENRFGFPCCVSKPLVAKOPFGFPDAPDKPVED 8416
Qy 6259 -----RGAPAL-----QETG----- 6268
Db 8417 VTSNSMLVKNWEPKNGSPILGYWLEKREVNSTHWSRVNKSLLNALKANVDGLLEGLTV 8476
Qy 6269 -----SOPPVGTSEAPVPPVPPQLLHEGPEGEPEAIARACEWTPIRM 6314
Db 8477 FRVCAENAGPKFSPSPDKTAHDPISPGPPIFRVTDTSTTIEL-----EWEPPAFN 8531
Qy 6315 EGAAMPG-----ACTGELLMD-----VHSHVVRRETTQ-RTYTQAIPTHTA----- 6354
Db 8532 GGGEIVGYFVDKQLVGTNE--NSRCPTEKMIKVRYTVKEIREGADYKLRVASVNAAGEGP 8589
Qy 6355 -----RPSMQVTIE--DYQAOQTGTAQFAILEGDPQPSVWTKDSQVLV 6398
Db 8590 PGETQPVTVAEPOEPFAVELDSVKGGIQIMAGTKLURIIPAVVTGRPVPTKVTWKEGSEL- 8648
Qy 6399 DSTRLSQQQEGTYSYLVLRHVASKDAGVYTCLAQNTGGQVLCKAEILLVGGNPE-----D 6454
Db 8649 DKDRVVIDNVGTKSELI--KDALKRKHGRVYITATNSCGSKGFAAARVEVE--DVPGPVLD 8705
Qy 6455 SEKQSHRRK--LHSFVYEKEEIRGVFGVFKRVQHKGNKILCAAKFIPLSRTRAQAYRE 8512
Db 8706 LKPVVTRNRKMLNWSDDDDGGSEITGFI--IERKDAKN----- 8743
Qy 6513 RDILAALSHPLVTGLDQFETRKTLLILELCSSEBLLDRLYKGVVTEAEVKVYQOVLV 6572
Db 8744 -----HTWRQPIETERS-KCDITGLL 8763

QY 6573 EQLHYLHSHGVL-----HLDIKPSNIMLVHAPAREDIKIDFGFAQNIITPAELOFSQ 6623
DB 8764 EGQEY--KFRVIANKKFCGPPVEIGP--ILAVDP-----8794
QY 6624 YGSPFVSPFELIQNPVSEASDIWAMGVISVLSLTCSPFAGESDRATLLNVLEGRVSW 6683
DB 8795 LGPP--TSPE-----RLTYERTKST-----ITLDWKEPRNGG 8826
QY 6684 SPMAHLSE---DAKOFIKATLQAPQAPSAQCLSHPWFLKSPAEBAHFNITKOLK 6739
DB 8827 SPIQGYIIEKRHRKDPF-----ERVNKLCTPTSLVENLDEHQMYEPRVK--- 8873
QY 6740 FLLARSRQWSIMSKSLVMRESIPELLRGPPDSPSLGVARHLCREDTGGSSSSSSSDNE 6799
DB 8874 -----AVNEIGSEPSLP-LNVV-----IQDDE 8895
QY 6800 LAPPARAK-----SLP---PSFVTHSPL-----LHPRGFLRPSASL---PEE 6835
DB 8896 VPPTIKLRLSVRGDTIKVKAGEPVHVPADVTGLPMPKIENSKNETVIEKPTDALQITKEE 8955
QY 6836 AEASERSTEAPASPAGCPGAPPAQCVPRHSVIRSLFYHQAGESPEHGALAPGSRHP 6895
DB 8956 VRSRAKTELSIPKAVREDKDTYTVTASNRILGVSFRNVHVEYDR-----PSPPRNL 9007
QY 6896 ARRRHLLKGGYIAGALP---GLREPLMEHVRVLEEAAREEOA-----TLLAK----- 6939
DB 9008 AVTDIKAESCVLTWDAPLDNGSE--ITHYVIDKRDASRKKAEVEVNTAVEKRYGIWK 9065
QY 6940 -APSPETALRLPASGTHIAPHGSHLSHDSPTSPRPSSEACGEAQLRPSAPSGAPIRDM 6998
DB 9066 LIPNGQYFRVRAVVKY---GISDECKSDKVIQDP-----YRLFPGPP--GKP-KVL 9111
QY 6999 GHQOSKOLPST---GGHPTAQ--PERPSPDSP-WGQAPFCHPKQG-----SAPO- 7044
DB 9112 ARTKSMVSWTPPLDNGSGITGYWLEKREGSPYWSRABITKVGLKGVEFNVPRL 9171
QY 7045 -EGCS-PPHAPVAPCPGFPFGSCKEAPLVPSPP-FIGQFOAPPAPAKASPPPLDSKXGPG 7101
DB 9172 LEGVKYQPRAMAINAAGIPSE-----PSDEVAGDPIFFPGP--PSCPEVKDKT-KS 9222
QY 7102 DISLPGRPKPCSPGASQASSOVSLSLVGSSQVGTGTEPGSLDAEGM--TOEAEDLS 7159
DB 9223 SISLGNKP---PAKDG---SPKGYIVNQEEGT-----TDWGRVNEPKLI 9264
QY 7160 DS---TPTLQRPQEQATMRKXFLSGRGYAGVAGYGTFAFGDAGMLOGGPNWARIW 7215
DB 9265 TTCECVNLEK-----LRKYP-----RVK- 9285
QY 7216 AVSQSEEEQEARAESQEQEARABESPLQVSARPVFVG-----RAPT 7262
DB 9286 AVNEAGESPDDTGTGEPATDQEE-----PEVFI---DIGAQDCLVCKASQIRIPA 9335
QY 7263 RSSPEPTP---WEDIGQVSLVQ---IRDLSDAADAADTISIDISEVDPAYNLNLDYDI- 7315
DB 9336 VIKGPTKSWEPDGRKAKAMKQGVHDIPEDAQLEAENSIVIIIECKRSHGKYSIT 9395
QY 7316 -KYLFFEMIRKVPKSAQPPSPGMAEELAEFP-EPTWMPGELG-----7360
DB 9396 AKNAGOKTANCRVKVMDVPFGPKDLKVSITRGSCLRSKMKMPDDGGDRKIGYVIEKRT 9455
QY 7361 -----PHAGL-----EITESEEDVDALLAAAGVKRKSPPSLSLFFPGRHL 7404
DB 9456 IDGKANTKVPDCSTTFVFDLISEQQYFRVRAENRFGIGPVFVIQNTARDP---IY 9513
QY 7405 PLDSPAELGLRERVKASY-----EHSIRLKGPEGLEKEG-----7440
DB 9514 PPDPPKILGLITKNTVHLSWKPKNDGGSPVTHYIVECLAWDPTGTKEAWRQCNRD 9573
QY 7441 -----PPRKPCLASFRLSGLKSM-----DRAPTF-LRELSDET 7474
DB 9574 VEELQFTVEDLVEGEYEFBRVAVNAAGVSPSATVFCQCRQDMPSPSIDLKEFME--V 9631
QY 7475 VLGSVTLACQVSAQPAQAATWSKDGAPLESSRVLISATLKNFOLL--TILVVVAB--- 7529

DB 9632 ESGTNVNVIAKIRGVFPFPTLTWFKAPPKPKDKPEVLVDTHVKNLVDDTCTLVIPQRR 9691
QY 7530 -DLGVYTCVSNALGTVTITGVLRKAERPSSPCPDIG-----EYVADGVLLWKPVESY 7583
DB 9692 SDTGLYTTAVNNLCTASKEMELNLVGRFG---PPVGPVKFESVSAQDMTSLNFPKDD 9747
QY 7584 GP---VTIVQV-CSLEGGSWTTLASDIFDCCVLSKLSRGGTYTFTACVSKAGNG-PYS 7638
DB 9748 GSKITNYVIEREANKRTVHVSVSEPKCTYITPKLLEHGYVFRIMAQNKGIGEPUD 9807
QY 7639 SPSE 7642
DB 9808 SEPE 9811
RESULT 8
Q10466 PRELIMINARY; PRT; 26926 AA.
AC Q10466;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Titin, heart isoform N2-B (EC 2.7.1.-) (Connectin).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE=96026330; PubMed=7569978;
RA Label S.; Kolmer B.;
RT "Titins: giant proteins in charge of muscle ultrastructure and
RT elasticity.";
RL Science 270:293-296(1995).
RN [2]
RP SEQUENCE OF 22277-25376 FROM N.A.
RX MEDLINE=9225380; PubMed=1582406;
RA Label S.; Gautel M.; Lakey A.; Trinick J.;
RT "Towards a molecular understanding of titin.";
RL EMBO J. 11:1711-1716(1992).
RN [3]
RP SEQUENCE OF 1976-2014 FROM N.A.
RA Label S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95331314; PubMed=7607248;
RA Gautel M.; Castiglione-Morelli M.A.; Pfuhl M.; Motta A.; Pastore A.;
RT "A calmodulin-binding sequence in the C-terminus of human cardiac
RT titin kinase.";
RL Eur. J. Biochem. 230:752-759(1995).
RN [5]
RP FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
CC ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
CC SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
CC KINASES.
CC -!- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112
CC IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE
CC DOMAINS.
DB EMBL; X64698; CAA45939.1; -
DB EMBL; X83270; CAA58243.1; -
DB EMBL; X64697; CAA45938.1; -
DB EMBL; X30568; CAA62188.1; -
DB EMBL; X64699; CAA45940.1; -
DB PIR; I38344; I38344.
DB PDB; 1BPV; 12-AUG-99.
DB PDB; 1GIC; 12-OCT-01.
DB PDB; 1NCT; 08-NOV-96.
DB PDB; 1NCU; 08-NOV-96.
DB PDB; 1TIT; 11-JUL-96.

DR PDB; 1TTU; 11-JUL-96.
DR GO; 0005856; C:cytoskeleton; IEA.
DR GO; 0016020; C:membrane; IEA.
DR GO; 0005524; P:ATP binding; IEA.
DR GO; 0005516; F:calmodulin binding; IEA.
DR GO; 0004896; F:hematopoietin/interferon-class (D200-domain. . .); IEA.
DR GO; 0004601; F:peroxidase activity; IEA.
DR GO; 0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; 0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; 0005198; F:structural molecule activity; IEA.
DR GO; 0005975; P:carbohydrate metabolism; IEA.
DR GO; 0007511; F:muscle development; IEA.
DR GO; 0006468; F:protein amino acid phosphorylation; IEA.
DR GO; 0006979; F:response to oxidative stress; IEA.
DR InterPro; IPR000382; Cytok receptor_2.
DR InterPro; IPR000577; FGGY Kin.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN-III-like.
DR InterPro; IPR001092; HLH Basic.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR008266; Tyr_kinase_AS.
DR Pfam; PF00041; fn3; 132.
DR Pfam; PF00047; ig; 91.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00060; FN3; 127.
DR SMART; SM00408; IGc2; 23.
DR PROSITE; PS00933; FGGY_KINASES_1; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS50835; IG LIKE; 87.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;
KW Serine/threonine-protein kinase; Alternative splicing; Repeat;
KW Immunoglobulin domain; Phosphorylation.
FT DOMAIN 1370 1389
FT DOMAIN 4429 4614
FT DOMAIN 24731 25070
FT DOMAIN 25030 25056
FT MOD_RES 1372 1372
FT MOD_RES 1377 1377
FT MOD_RES 1382 1382
FT MOD_RES 1387 1387
FT MOD_RES 26171 26171
FT MOD_RES 26178 26178
FT MOD_RES 26184 26184
FT MOD_RES 26190 26190
FT MOD_RES 22277 22277
FT CONFLICT 22449 22449
FT CONFLICT 22454 22454
FT CONFLICT 23324 23324
FT CONFLICT 26326 AA; 2993428 NW; DSEECDS254DF5523 CRC64;
FT SEQUENCE
Query Match 8.7%; Score 3577.5; DB 4; Length 26926;
Best Local Similarity 20.2%; Pred. No. 4.6e-181;
Matches 1929; Conservative 1288; Mismatches 3662; Indels 2663; Gaps 320;
QY 4 POFSGAPRLTPKPAFVSVGKDALISQIVGNPTPQVSWKQDPVTAGAFRLA---Q 60
DB 1030 PGEAPAPFTRPVQKLVEGSGVVGQVGNPKPXYVWKKSGVPLTTGYRYKVSYNKQ 1089
QY 61 DGDLYRLTLDLALGDSGYVCARNALCEFAAVGL-----QVD 100
DB 1090 TGECC-KLVISWTFADAGEYTIWVKNKEGTSASASALLEADYELLMKSSQEMLYQVQT 1148

QY 101 A---EAACAEQAPHF----- 113
DB 1149 AFQPEPEVGTATGFFVSEYEKEQALIKKMAKDTVVVTVYVEDQEFHISSEERL 1208
QY 114 -----LRPTSIRVREGSEATPCRV 133
DB 1209 IKEIYRIIITLLEELLEDGEEKMAVDISESAVSGFDLRINKYRIILEGMGVTHCKM 1268
QY 134 GGSPPRAVSWSKGRRLGEPDGPVRVVEELGEASA-LRRAAPRDPGGTYVEVAENPLGA 192
DB 1269 SGYPLPKIANYKDGKII--KHGERYQWDFLODGRASLRIPVLPPEDEGIYAFASNIKGN 1326
QY 193 ASAAALVVDSDAADTASR--POTSTAALLAHLQRR--REAWRASGAPASPP----- 240
DB 1327 AICSGKLYVEFAAPLGAFTYIPTLEPVSIRISLSPRSVSRSPIRMSPARMSPARM 1386
QY 241 -----STGTR-----TCTVTEGKHARLSCVVTGEPKPTWKK 273
DB 1387 SPARMSPPGRLEETDESQLERLYKPVVLKPVSKLEGANGCRFDLKVVGRPMPTFWFH 1446
QY 274 DGQLVTEGRRH--VYVEDAQENFVLKILFCQSDRGLYTCTASNLVGYQTVSSVLVVV-- 328
DB 1447 DGQIIVNDYTHKVIKEDGTQSLI--IVPATPSDGEWTWVAQNAGRSSISVILTVEAV 1504
QY 329 ----- 328
DB 1505 EHQVPMFVEKLKNVNIKEGSRLEMKVRATGNPNPDIVLKNSDIIVPHKYPIRIEGTK 1564
QY 329 -----R 329
DB 1565 GEAAALKIDSTVSQDSAWYTATINKAGDTRTCNVNVEVEFAPEPEPERKLIIPRGYRAK 1624
QY 330 EPAVP-----FKRLQDLEVRKESEATFCE---VP 357
DB 1625 EIAAPELEPLHLRYGOEWBEGDLYDKQKQFFKKKLTSLRLKRGFAHFCRLTPIS 1684
QY 358 QPSTEAAPKEETRL-----WASAKYGI-----EEEGTERRLT 390
DB 1685 DPTVWEHLHDGKPLEAANRLMINEFGVCSLDYGVAYSRDSGIITCRATNKYGTHTSA 1744
QY 391 VRNVSADDDAVYICETPEGSR-----TVAELAVQGNL-----LRKLPRKT 430
DB 1745 TLIVKDEKSLVESQLPEGRKGLQRIELEMAHEGALGVTTDQXKQKPDVILYPEPV 1804
QY 431 AVRVGDTAMP-CVELAVPVPVHMLNQSEVAVAGVAISAEGRHTLTISQCLSDVQ 489
DB 1805 RVLEGETARPCRVTVGPQPKVNNYLNGLIRKSKFRVRYDGI-HYLDIVDCKSYDTGE 1863
QY 490 VAFWAG-----DCQTSTRECVSAPRKPPLQ-----FVD 518
DB 1864 VKVTAEINPEGVIEHKVLEIQQREDFRSLRAPEPRPEFHVHEPGKLFQEVQKVRPVD 1923
QY 519 -----PVVKARMSSVILSWSPPHGE-----RPVTIDGY-----LVEKKLGTW 560
DB 1924 TTETKEVVKLRAERITHEKVPSEBELSKFKRRTEEGYEAITAVELKSKKDSYEE 1983
QY 561 I-----RCHAEWVATPELTVD---VAEENGFQFVSALNSFGQSPYLEPFGVHLAPKL 613
DB 1984 LLRKTKDELLHW--TKELTEEEKKALAESEKTIPTFKPKDKIELSFSME-----APKI 2034
QY 614 AVTPLKAVQAVGEGVTVFSDLTVASAG-----EWFLDGQALKASS-VYEIHCORTRH 666
DB 2035 FERIQSQTVG--QGSDAHPRVRV---VGKPOPECEWYKNGVKIERSDRIYWPEDNVC 2088
QY 667 TLITIREVPASLHGAQLKF-----VANGIESIRMEVTAAPGLTANKPAPAAAREVLARLHE 722
DB 2089 ELVIRDVTAE-DSASIMVKAINTAGTSSHAFLVQAKQLITFTQ-----ELQDVVAKEXD 2143
QY 723 E-AQLLAEISDQAAAATWLKDGFTLSFGPKYVQASAGRRVLLVRDVARDDAGLVECV-- 779
DB 2144 TMAFTFECTSEPFVKVKKVKGMEVHEGDKYRHMHSRDKVHFLSILITDTSDAEDYSCVLV 2203

Db 4271 EVNWTPEKLVPSDEKFKCLODQNTYTLVLDKXNTEDHQGEYVCEALNDS-----GKTAT 4325
Qy 2459 SARL-VVSDIPVILTRPLEKPTGRELQSVLWSCDPRPAPKA-VOMYKDDTPLSPSEKPKM 2516
Db 4326 SAKLTVKRAAPVIRKRIEPLVALGHLAKFTCBISQAPNVRVQWFRAGRIEYSDKCSI 4385
Qy 2517 SLEGOMAEIRLILMPADAGVYRQAGSAHSSTETVVEAREVTVTG-----PLODA 2567
Db 4386 RSSKYIISLEILRTQVDCGYTCKASNEYGSVCTA-TLTVTPGGEKKVRKULLPERKP 4444
Qy 2568 EATEGMAFSCELSHDEEVEWLSNGMPLNDSFHEISHGRHRTLVLKSIQRADAGIV 2627
Db 4445 EPKEE--VVLKSVLRKRPEEPEKPKL-----EKVKKPAVPEPPPPKPVBEVEPTV 4497
Qy 2628 RASSLKVSTARSARLEVRVVPVFLKALDLSABERGTLALQCEVSDPE-----AHV 2678
Db 4498 TKRERKIPEPTKVP-EIKPAFLPAPEPKPKEAEVTKTKPPPEPPTPIAABVTPVV 4556
Qy 2679 WRKQVQLGSPDKVDFLHTAG-TRGLVVHDVSPDAGLYTCHVGS-----EETRARVRV 2731
Db 4557 GKABAK--APKEAAKPKPIKGVPKTSPIEAERKLRPGSGGKPPDEAPFTYQL 4613
Qy 2732 HDLHVIGITKRLKTMVLE-----GBSCSPCEVLSHESADPANWTVGKTVGSSSRFOATR 2787
Db 4614 KAVPLKFKYKEIKDIILTESEFVGSSAIFECVLSPSTAI--TWMKDGNSIRESKHFPIA 4671
Qy 2788 QGRKYLIVREAAAPSDAGEVVFVSR-----GLTSKASLIVRERPAAIKPLEDO-WVAPGE 2842
Db 4672 DKORKLHIIIDVLSDAGEYTCVLRLNGKEKSTAKLVEELPVRFVKLTLEEVTVVKGQ 4731
Qy 2843 DVELCELSRAGTPVHMLKDRK-AIRKSKYDVWCEGTMAMLVIRGASLKADAGTYCEVE 2901
Db 4732 PLYLSCELNKE-RDVVRKDGKIVVEKPGRIVPGVIGLMRALTINDADTDAGTYTVTVE 4790
Qy 2902 ASK-----STASHVEEKANCETEELTNLOVEKGTAVFTKTEHPAAVTVRKGLLELRA 2957
Db 4791 NANNLECSCKVKEVIRDLVKPIRDGHVXPKGTAFACDIAKDTPNIKMFKGYDEIPA 4850
Qy 2958 --SGKHQPSQEGTLRLTALSLEKADSDTYTCIDGQAQSRALVVOGRVHIIIDLEDVD 3015
Db 4851 EPNDKTEILRDGNHLYLXIKNAMPEDIAEYAVEIEGKVPKAKLTJGEREVELLPKIEDVT 4910
Qy 3016 VOEGSSATFRCHISFANVEPVHFLDKTPLHANELNEIDAPGGVHVLTLRLQALKOSGT 3075
Db 4911 IYKESASFDAISEADI-PGOMKLKGELLRSPPTCEIKAB-GGKRFTLTHKVLDDQAGE 4968
Qy 3076 IYFEAGDQASAAARVTEKPSVFSRELTDATITEGEDLTLCETSTCDIPMCWTQDKGTL 3135
Db 4969 VLYQALNALTALITVKIEILDFVPLKDVTPERRQARFEC-VLTREANVWSKGPDI 5027
Qy 3136 RGSARCQLSHEGHRACQLITGATLQDSGRYKCEAGGACSSIRVVRHAPVRFQBALXDL 3195
Db 5028 XSSDKFDIADGKHILVINDSQFDEGVYTAEEVEGKTSARLFVTGIRLAFMSPLEDQT 5087
Qy 3196 VLEGGAATLRCVLSVAAPVKCYGNVLRPGDKYSLRQEGAMLELVNLRPODSGRYS 3255
Db 5088 VKEGETATFVCELSHEKHMVWFKNDAKLHSTRTVLISSEBTKHLEMKETVLDDISOIK 5147
Qy 3256 CSFGDQTTSATITVATPAQF-GKLRNKEATEGATATLRCLESTKAPVWRKGSSETLRDG 3315
Db 5148 AQVKELSSTAQLKVLLEADPVYTKLHDKTAVEKDBITLKCEYSKDVVPVKWFGDBEIVPS 5207
Qy 3316 DRYCLRODAMCELQIRGLAMVDAEYSCVCEERTSALTRPMPAHFIFGLRHQESIE 3375
Db 5208 PRYSIKADGRILKIKKADLKDGEYVCDGCTDKTKANVTVEARLIEVEXPLYGVEVVF 5267
Qy 3376 GATATLRCELS-----KAAPVWRKKGRESLRDGRHSLRQDAGVCELQICGLAVA 3425
Db 5268 GETAHEIELSFPDVHGQWKLKQPLTASPDCEIIEDGKX-----ILLHNLQGLM- 5319
Qy 3426 DAGEYSCVCGEBRTSATLTKVALPAKFTTEGLRNEEAVEGATAMLCESKVAAPV-EWRKG 3484
Db 5320 -TGEVSFOANAKSAANLAKVELPLTIFITPLSDVKVFEKDEAKFECEVSREP KTFRLKG 5378

Qy 3485 PENLRDGRYILROBGRCELOICGLAMADAGEYLVCVQGBERTSATLIRALPARFIEDV 3544
Db 5379 TOBITGDDRFELIKDGTKHSVMIKSAAFEDAKYMFEAEDKHTSKGLIEGRILKPLTFL 5438
Qy 3545 XNCEAREGATVLCQEL-NSAAPVWRKGSSETLRDGRYSLRQDGTCKCELOIRGLAMAT 3603
Db 5439 KDVTAKEKESAVFTVELSHDNLRVKNQDLHTTRSVMQDEBGTHTSITFKOLSIDDT 5498
Qy 3604 GEYSCVCGGBERTSAMLTVRALPIKFTTEGLRNEEATEGATVRLCELSKM-AVEWVKGE 3662
Db 5499 SQIRVEAMGVSBEAKTLVLEGDPYFTGKLDQYTGVEKDEVILQCEISKADAVKVKFKOK 5558
Qy 3663 TLRDGRHSLRQDAGARCELQIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPKFTBGLN 3722
Db 5559 EIKPSKNVITKDGKEMLILKALKSDIGQYTCOGTDTKSGKLDIEDREIKLVRPLHS 5618
Qy 3723 EE----- 3724
Db 5619 VEWNETARFETEISEDDIHANWKLKGEALLQTPDCEIKBEGKIHSVLHNCRLDQOG 5678
Qy 3725 -----ATEGDTATLWCELS-KAAPVWRKKGHEIL 3752
Db 5679 VDFQAAVNVKSSAHLRVKPRVIGLLRPLKDVTVTAGETATFDCLESDYEDIPVWYLGKGL 5738
Qy 3753 RGDHSLRQDGRCELOIRGLAVVDAGEYSCVCGGBERTSAMLTVRALPARFIEDVKNQE 3812
Db 5739 EPSDKVVRSEKGVHTLTLRDVKLEDAEGVQLTAKDFKTHANLKVPEPVEFTKPLEDQT 5798
Qy 3813 ARGATAVLQCELSKA-APVWRKGSSETLRDGRYSLRQDGTCELOIRGLHGLSVADTGEYS 3871
Db 5799 VEGATAVLECEVSRENAKVKMFKNGTEILKSKYIEIVADGRVRLKVLINDCTPEDTKYT 5858
Qy 3872 CVCGGBERTSAMLTVRAPQPVREPLQSLQAEGBSTATLOCELSEPTATVWVSGGLQLOA 3931
Db 5859 CDAKDFKTSCLNVVPHVEFLPLDQLVQREKEMARFECLESRENAKVKWFKDGAIEIK 5918
Qy 3932 NGRREPLQCELAELVLQDREDTGEYTCOSQATSATLTVTAAPVRFELRQEQEVD 3991
Db 5919 GKDYIISKGAIRLVINKCLLDDEAEYSEVTAETSGMLTVLEBEAVFTKLANIEVS 5978
Qy 3992 EGCTAHLCCLESRAGASVWRKGSLOLFPCKAYQVQDGAALVVRGVEQEDAGDYTCD 4051
Db 5979 ETTIKLVCEVSPGAEVWYKGBEIIETGRYEIITBGRKILTVIONAHLEDAAGNYNCR 6038
Qy 4052 TGTQSMASLSVRVPKPKTRQLQSLQEQETDIARLCCOLSDAESGAANVOWLKEGVELHA 4111
Db 6039 LPSRTDGVKVKHLEAAEFISKPNLEILEGEKAEFVCSIS--KESFPVQWKEDDKTLES 6096
Qy 4112 GPKYVRSQATRELLIHOLEAKDTGEYACVTGGQXTAASLRVTEPEVTVIRGLVDAEVT 4171
Db 6097 GDKYDVIADGKRVLVVKOATLQDMGTYYVWGAARAAHLTVIE-KLIRVPLKDRVK 6155
Qy 4172 ADEDFEFSCEVSRAGATGVQWCLQGLPLOSNEVTEVAVRDGRIHTLRLKGVTPEDAGTVS 4231
Db 6156 EQGEVFNCEVNTGAKAKWFRNEEAI FDSKY--IILQKDLVYTLIRDAHLDDQANYN 6213
Qy 4232 FHLGNH-----ASSAQLTVRAPVETILEPLQDVLQSEGQDASQCLSRASGOEAWLG 4286
Db 6214 VSLTNHGRGNVKSAAANLIVVEEDLRIPELKDITETWEKKSVTFWCKVNRNL-VTLKWTKN 6272
Qy 4287 GVPLQANEMDITVEOGTTLHLTHKVTLEDAGTVSFHVGTCSSEAKL-VTAKNTVVRG 4345
Db 6273 GEEVFPDNRVSYRVDK-YKHMJTIKDCGPDGEYIVTAGQDSKVAELIIEAPTEFVEH 6331
Qy 4346 LENVEALEGGALEFECOLSOPEVAANTWLLDDBPVTSENAEEVFPENGRLHLLILKNLR 4405
Db 6332 LEDQTVTFDDAVFSQOLSR-EKANYKVRNGREIK--EGKTKFKDGSIHLLIIXDCR 6388
Qy 4406 PQDSCTVTFLAGDMVTSARLTVRGWRLILEPLKNAAVRAGAQAARCTCTLSEAVPVEAS 4465
Db 6389 LDDECEYACGVEDKSRARLVEEIPVPIRPPQDILEAPGADVFLAELNK--DKVEVQ 6446

QY 4466 WYINGAAVQPDSDWTVTADGSHOALLRSQAOPHHAGBVTACRDVASAR----- 4516
Db 6447 WLRNNMVVQGDKH-QMMSEKIHRLQICDIKPRDQGEYRFAIAKXKARAKLEAAAPKI 6505
QY 4517 ----- 4516
Db 6506 KTADQDLVVDGKPLTWVPYDAYPKAEAEWFKENEPLSTKTIDTAEQTSFRILEAKKG 6565
QY 4517 -----LTVLGLPDPEDAEVVAHSHHTVLSWAAPMSDGGGLC 4555
Db 6566 DKGRYKIVLQNHKGKAEFINLKVIDVPGVVRNLEVTETFDGEVSLAWEELPTDGGSKII 6625
QY 4556 GYRVEKGCATQWRILCHELVGPECVVDGLAPGET-YRFRVAAGVPVGAEPVHLPTQV 4614
Db 6626 GYVVERRDIKRTWLTADRAESCEFTVTGLQKGVYLFVRSARNRVGTGEPVETDNPV 6685
QY 4615 RLAEPPKVPPOP-----SAPESRQVAGEDVSLEVVABAGEVIMHKGM 4660
Db 6686 E-ARSKYDVPGLNVTITDVRFGVSLTWEPPEYDGAETNTYVIELRDKTISRMDTAM 6744
QY 4661 ERIQPGGRFVVYSGROQMLVIKGFATBEDQGEYHCGLAQGSICPAATFOVALSPASVDE 4720
Db 6745 T-----VRAEDLSATVDVSGQYSFRVRAQNRIGVK-----PSAATPFVKVADPIERP 6795
QY 4721 APQSLPPEAAQEGDLHLWLALARKMSREPTLDSISELPEDGRSQRLPQBAEEVAP 4780
Db 6796 SPPVNLTSDDQTQSSVQLKWBPLKD---GGSPILGVIIERCEE-GKDNWIRCNM-KLVP 6850
QY 4781 DLSEGYSTADELARTGDADLSHTSDDESRAGTSLSVTLKKGACRPGTSLASKVGAPAA 4840
Db 6851 ELTYKVTGLEK-GNKYLVRSAENKAGVSDSEI----- 6883
QY 4841 PSVKPQOOQELAAVRPPLGLDSTKD-LGDPMSMDKAAVK-----IOAAFKGYK 4887
Db 6884 -----LGPLTADDAFVPTMDLSAFKDGLEIVPNPITILVPSTGY- 6924
QY 4888 VRKEMKQEGPMFSHTFGDTAQVGDALRLCECVASADVARWKLDCVBLTDCRHHIID 4947
Db 6925 -----PRPTATWCFDGVLETDGDRVMKTLA----- 6951
QY 4948 QLGDCSTLLIAGLDRAAGCYTCQVSN-----KFGQVTHSA 4984
Db 6952 -----YAEVLISPSERSDKGIYTLKLENRYKTIISGEIDVNVVAPSPAPKELKFGDITKDS 7006
QY 4985 CVVSGSESEABSSGGLDDAFRRARLRLHRLFRK----- 5021
Db 7007 VHLT-----WEPDDGGSPLTGYVVEKREVSRTKWTVMDFVTDLEFTVPLDVGKEYLF 7062
QY 5022 -----SPAESDEELFLSADGCPAEPPEPADWQTYREDEHI----- 5058
Db 7063 KVCARNKCGPEPAV-DEPNVMTPTATVPDPPENVRKWRTRANSIFLTDPPNQGGSR 7121
QY 5059 -----CIR-----FPALEARQAVTRFOE--MFATLIGVBEIKLVEQPRRVM 5100
Db 7122 IKGYIVERCPRGSKWACGEPVAETKVEVTGLEEGKWA-----YRKVT 7166
QY 5101 CISKETAPVPPPLSLLTSDRAAPFLATE--LQNOEVQDGPVPSFDCVTCQPPSVR 5158
Db 7167 LNRGASKPSRPTTEIQAQVDTQEAPEIFLDVLLAGLTVKAGTKIELPATVTKRPEKIT 7226
QY 5159 WFKDGKLEEDHYMINEDQGGHOLIITAVVPADMGVYRCLAENSGVSTKAEELRVDL 5218
Db 7227 WTKADMILKQKRITI-ENVPKSTVTIIVDSKRSDTGTIIIEAVNVCGRATAVVEVNV-L 7284
QY 5219 TSTDYDTAADATESSSYFSAQGYLSSREQSGTSTTDEGQLPQVVEZLRDLQVAPTRLA 5278
Db 7285 DKPGPPAFTDVTNESCILLTMNPPRDGGS----- 7316
QY 5279 KFOIKVKGYPAPRLYWFKDGQPLTASAHIMTQKILHLEIISVTRDSGQVAAVISA 5338
Db 7317 ----KITINYVER-----RATDSEVWHKL---SSTVKOINFKATKLIPN 7353
QY 5339 MGAAYSSARLLVRGPDPEEKPSADSVHEQLVPPRMLERFTPKVKKGS-SITFSVKVE-- 5395

Db 7354 KEYIFRVAENMYGAGEPVQASPIAKYQDFDPPGPPTRLPESDITDKDAVTLTWCBSDDDG 7413
QY 5396 GRPVPVTVHVLREBEAERGVLWIGPDTPGYTVASSAQOHSILVLDVGRQHGTY--TCIASN 5453
Db 7414 GSPI-TGYVW-----LDPDTDKVRCKMPVKDITYRVKGLTNKKKYRFRVLAEN 7464
QY 5454 AAOQALCSASLHVSGLP-KVEQEKVKKEALISTFLOG--TTOAI SAQGLFASPADLGGQ 5510
Db 7465 LAG-----PGKPSKSTEPILIKDFIDPPWPFGPKTVKDVGKTSVR-----LAWT 7508
QY 5511 RKEEPLAAKEALGHLSLAEVGTETEEFLQKLTSQITEMVSAKITQAKLQVPG-GDSDEDSKT 5569
Db 7509 KEHDGGAKEISYVIMELKTGTDEWR-----VAEGVPTTQHLPLGLMEGEYSFR 7559
QY 5570 PSASPHGRSRPSSSIQSSSESDGDARGEIDFYVVTADYLPGLGASQDAILTBREGY- 5628
Db 7560 VRAVNKAGESEPE-----PSDPVLCREKLYP 7586
QY 5629 -----VEVLDAAH--PLRWLVRTKP-----TKSSPSRQGW--VSPAYLDRRLK 5667
Db 7587 PEPPLWLEVINIKNTADLKWTVPEKDGSPITNIVIEKRDVRRKGWQTVDTTVKDTKCT 7646
QY 5668 LSP--EWG-----AAEAPFPEGEAVSBDEYKARLSSVIOELLSSQAFVEELQFLOSHH 5719
Db 7647 VTPLTEGSLYFRVAEE-----NAIGQSDY-----TEIEDSLAKDTFT----- 7685
QY 5720 LQHLERCHVPIAVAGKAVIFRNVDRDIGRHSSFLQELQCCDITDDVAMCFI-KNOAAF 5778
Db 7686 -----TGPPIYALA-----VVDTKRHVDLKWEPKNDGGRPIQRVVIKKEKELG 7730
QY 5779 EQYLEFLVGRV-----QAESVVVSTAIQEFYKAYE-----EALLAGDPSQ 5819
Db 7731 TRWV--AGKTAGPCNFRVTDVIBGTEVQ--FQVRAENAGVGHSPSEPTILSTEDPTS 7786
QY 5820 PPPPLQHYLE-----QOVER-----VORVQALLKELIRKA 5851
Db 7787 PPSPLDLHVTDAGRKHIAIAWKPEKNGSGSPITGYHVEMCPVGTETKRMVNSRPIKDLK 7846
QY 5852 RNRQNCALLEQAYAVVSALPORAENKLHVS-----LMENYPG-----TLEALGEPIRQ 5900
Db 7847 FKVESGVVPDKEY-----VLRVRAVNAIGVSPSEISENVAKDPCKRTID-----LETH 7897
QY 5901 HPIVNEG-----APGARMPWKGNHRHV-----FLFRN-----HLVICKP-RRDS 5938
Db 7898 DIIVTEGEKLSIPVPFRVAVPVTVSMHKGKGVKASDRLTKNDHISAHLEVPKSVRADA 7957
QY 5939 RDTVTYVFRNMKLSLIDLANDQVEG-----DDRAFEV-----WQ----- 5973
Db 7958 GIYTTI--LENKLSATASINVKVIGLPGCKDIKASDITKSSCKLTWEPPEFGGTPIL 8015
QY 5974 -----EREDSVRYLLQARTAIKGSW-VKEIC-----GIQORLALPV- 6010
Db 8016 HVLREERAGARTYIPVMSGENKLSWTVKDLIPNGEYFRVKA VKNVGGGSEYIELKNPVI 8075
QY 6011 ---WEPD--FEELADCTAELGETVKLACRVGTGTPKFSVSWYKDGKAVQVDHILLIE 6064
Db 8076 AQDPKQPPDPVDVEVHNPTAE-----AMTITWKPLP--YDGSKIM--GYIIEK 8121
QY 6065 DFDGS-----CALILDSLTVDSG--QYMCFAASAAGNCSTLGIKLIVQVPP-RF 6110
Db 8122 TAKGEERWKRCNEHLVPLITVTYKLEBEGKEYQFVRVAENAAAG--ISEPSRATPTPKA 8177
QY 6111 VNKVPAASFV-----EGEDAQFTCTIEGAPYQIRWYKD-----GALLTT 6150
Db 8178 VDPIDAPKVIILRTSLEVKRGEIDALDASISGSPYTTITWKNENVVPEETIKGAAPLVR 8237
QY 6151 GNKFO-----TLSEPRSGLLVLVIRAAASKEDGLGYCEGLVNLRLSARASA 6195
Db 8238 RKGEVQEEEPFVPLTQRLSIDNSKKGESQLRVDSILRDPHGLYMIKVENDHGIAPC 8297
QY 6196 ELRTQS-----PMLQAQEOCHREQLVAVE-----DTTLERADQZ-----VT 6232

Db 8298 TVSVLDTPGPPINFVFDIRKTSVLCKWEPPLDDGSEIINYITLKKDKTKPDSEWIVVT 8357
Qy 6233 SVLK-----RLGPK-----APGCS-----TGDLTGPGPCP----- 6258
Db 8358 STLRCHKSVTKLIEGKEYLFRVRAENRFGPPPCVSKPLVAKOPFGPPDAPDKPIVEDV 8417
Qy 6259-----RGAPAL----- 6264
Db 8418 TSNMLVKNPEKONGSPILGVLKREVNSTHWRVKNKSLNALKANVDGLLEGLTYVF 8477
Qy 6265---QSTGQPPVTGSEAPVPRVPOPLLHGEQPEPEAJARAOEWTPVIRMGAAWPG 6321
Db 8478 RYCAENAAAGP---GKSPSPDPTAHDPLSPGP---PIPRVTDSTSTIELEWPPAFNG 8532
Qy 6322 AG-----TCELLWD-----VSHVIVREITQ-RTYTYQAIDHTA----- 6354
Db 8533 GGEIVGYFVDKOLVGNKWSRCTEKVIVQVTVKXIEGADYKLRVSAVNAAGEGPPGE 8592
Qy 6355-----RPPSMQVITIE---DVQAOTGTAQPEAIIIEGDPQPSVTVYKDSVQLVDST 6401
Db 8593 TQPTVAVPEQPEPAVELDVSKGGIIMAGKTLRIPAVVTGRTVPTKWTKEEGL-DKD 8651
Qy 6402 RLSQOQEGTYSVLVRHASKDAGVYTCIAQNTGGQVLCHELLVLGGDNPE---DSEK 6457
Db 8652 RYVIDNVGTSKSLIIXDALRKHGRYVITATNSCGSKFAAARVEVF---DVPGPVLDPK 8708
Qy 6458 QSHRRK-LHSFVEKKEIGRGVFGVKVQHGKNIKILCAAKFIPLRSRTAQAVRREDI 6515
Db 8709 VVTRKMLLANSDDDDGSGSITGFI---IERKDKA----- 8743
Qy 6516 LAALSHPLVTGLDQFETKTLILILELCSSELDRLRYKGVVTEARVKYIOLVEGL 6575
Db 8744-----RGAPAL----- 6264
Qy 6576 HYLHSHGVL-----HLDIKPSNIMVHAREDIKIDGFGPAQNIPTAELQSOYGS 6626
Db 8767 EV--KFRVAKNKGCGPPVEIGP--ILAVDP-----LGP 8797
Qy 6627 PBEVSPETIQONPVSEASDIWANGVISVLSLTCSSPFAGESDRATL-LNVLEGRVSWSSP 6685
Db 8798 P--TSPE-----RLTYT-----ERQSTITLDWKEPRNGGSP 8828
Qy 6686 MAHLSE---DAKPIKATLORAPQAPSAACLSHPWFLKSMAPAEAHINTKQLXEL 6741
Db 8829 IQGYIEKRRHDKPDF-----ERNKRLCTPTFLVENLDEHQWYFRVK----- 8873
Qy 6742 LARSQRSLMSYKSLVMSRIPELLRGPDPSPGLVARHLCRDTGGSSSSSSSDNELA 6801
Db 8874-----AYNEIGSEPSLP-LNV-----IQDVEP 8897
Qy 6802 PFARAK-----SLP---PSPVTHSLP-----LHPRGFLRPSASL---PEAE 6837
Db 8898 PTIKRLSVRGDTIKVKGEPVHIPADVTGLPMPKIEMSKNETVIEKPTDALQITKEVS 8957
Qy 6838 ASERSTEAPAPASPEGAGPPAAQGVPRHSVIRSLFYHQAGESPEHGALAPGSRHHPAR 6897
Db 8958 RSEAKTELSIPKAVREDKGTYYVTASNLGSRVFRNVHVEYDR-----PSPPRNLAV 9009
Qy 6898 RRHLLKGGYIAGALP---GLREPLMEHRYLVEEAREEOA-----TLLAK-----A 6940
Db 9010 TDIIKAESCYLTWDAPLDNGSE--ITHYVIDKRDSARKKAEMEVTNTAVEKRYGIWKLI 9067
Qy 6941 PSFETALRIPASGTHLAPHSHLSLEDSFSTRPSPSEACGEAQLRPSAPSGAPIEDMGH 7000
Db 9068 PNGQYFRVRANKY---GISDECKSDKVIQDP-----YRLPGPP--GKP-KVLAR 9113
Qy 7001 PQGSKOLPST-----GCHPQTAQ--PERPSPDGP-WGQAPFCHPKQG-----SAPO--E 7045
Db 9114 TKGSMVSWTTPPLDNGSGSITGWLEKREESGYNSRVSRAPITKVLKGVEFNVRLLLE 9173
Qy 7046 GCS-PHPAVAPCPGPGSPGCSKEAPLVPSSP-FLGQOAPAPAKASPLPDSKMGPGDI 7103
Db 9174 GVKYQFRAMAINAAGICPSP-----PSDPEVAGDPIFPFGP-PSCPEVKDKT-KSSI 9224

Qy 7104 SLPRKFPGPCSPFGSQAQSSQVSLRVGSSQVCTEPGPSILDAEGW--TQEAEDLSDS 7161
Db 9225 SLGWKP---PAKDG---SPIKGYIVEMQEGT-----TDKRVNEPDKLIIT 9266
Qy 7162---TPTLQRPQOATMRKFSGLGGRGYAGVAGYGTFAFGDAGGMLGQGPMMARJAWAV 7217
Db 9267 CECVVPNLKE-----LRKYRF-----RVK-AV 9287
Qy 7218 SQSEEBEQEAREASOSEEQEAREASPLPOVSARPVPEVG-----RAPTRS 7264
Db 9288 NEAGESEPSDTTGEIPATDIQES-----PEVFI---DIGAQCILVCKAGSOIRPAVI 9337
Qy 7265 SPPEPTP---WEDIGQVSLVQ---IRDLSDGAEEAADTISLISEVDPAYNLNLSLDYDI--K 7316
Db 9338 KGPTPKSSWEPFGKAKKMGVHDIPEDAQLETAENSGSVIIIPCKRSHTKYSITAK 9397
Qy 7317 YLPEFEMIFKPKSAQPPSPMAEEELAEFP-EPTWPPGELG----- 7360
Db 9398 NKAGQKTANCRVKVMDVPGPPKDLKVSDITRGSCLSKWMPDDDDGDRINKGYVIEKRTID 9457
Qy 7361-----PHAGL-----EITESEDDVALLAAAVGRKRKWSPPSRSLFHPGGRHLPL 7406
Db 9458 GKAWTKVNPDCGTTTFVVDLLSEQQYFFRVRAENRFGIGPPVETIQTARTDP--LYPP 9515
Qy 7407 DEFAELGLRERKASV-----EHSIRILKGRPEGLEKEG----- 7440
Db 9516 DPPIKLKIGLITKNTVHLSWKPKPKNDGSPVTHIVECLAWDPTGTKEAWRQCNKRDVE 9575
Qy 7441-----PPRKPKGLASRLSLGKSW-----DRAPTF-LRELSDETVVL 7476
Db 9576 ELOFTVEDLVEGGEYFRVAVNAAGVSPSATVPCDCQORPMDPSPIDLUKEME---VEE 9633
Qy 7477 GQSVTLACQVSAQAQATWSKDGAPLESRRVLIISATLKNFOLL--TILVVVAE-----D 7530
Db 9634 GTNVNIVAKIKGVPPFTLTWFKAPPKPNKEPFLVDTHVNLVDDTCTLVIPOSSRSD 9693
Qy 7531 LGVITCSVALGTVTTTGVLRKAERSSSPCPDIG-----EYVADGVLLVWKEVSEYGP 7585
Db 9694 TGLYITIAVNLGTASKEMRLNVLGRPG---PVGPIKFESVSADQMTLSWFFPKDDGG 9749
Qy 7586---VTVYIWO--CSLEGGSWTTLASDIPCCYLTGKLSRGGTYTERTACVSKAGMG-PYSSP 7640
Db 9750 SKITNYVIERKRNKTVHVSSEPEKCTYIPKLLEGHEVYVFRMAQNKYGIGEPDSE 9809
Qy 7641 SE 7642
Db 9810 PE 9811

RESULT 9

Q8WZ42 ID Q8WZ42 PRELIMINARY; PRT; 34350 AA.
AC Q8WZ42;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE titin.
GN TTN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20309627; PubMed=10850961;
RA Freiburg A., Trombidas K., Hell W., Cazorla O., Fougereuse F.,
RA Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,
RA Granzier H., Labeit S.;
RT "Series of exon-skipping events in the elastic spring region of titin
as the structural basis for myofibrillar elastic diversity";
RL Circ. Res. 96:1114-1121(2000).
RN [2]

Db 4395 ALGHLAKFCEIQSAPNVRWFOWFKAGREIYESDKSIRSSKYISSILEILRTQVDCGEVT 4454
QY 1054 CSARGORVSF---RLHITE---PKMFAKEQSVHNEVQAEAGASAMLSCEVAQAQT-EVT 1106
Db 4455 CKASNEYGSVCTAILVTIEAAPTFLSPKSLTTFV---CKAAKFTCTVTGTPVIEITI 4510
QY 1107 WYKDGKLSSSKVGNEVKGCTRRLLVPOAGKADAGEYSCEA---GGCRVSFHLHITEP 1162
Db 4511 WOKDGAALSPSPNWKISDAENKHIELSNTIQRDGRVYSCASNKFGADICQAEILIID- 4569
QY 1163 KGVFAKEQSVHNEVQAEAGTAMLSCEVAQP-QTEVTWYKDGKLL--SSSSKVRNEVKG 1219
Db 4570 KPHFIKEL---EPVQSAINKVHLECOVDDESKVTWWSKDGCKLPPOKGYKICFEDKIA 4626
QY 1220 TRRLVVOQKADAGEYSC---EAGGORSVQLHITPFAKFAKEQLVHNEVTEAGAS 1275
Db 4627 TLEIPLAKL--KDSGTIVCTASNEAGSSCSATV-VREPPS-FVKK--VDPSPYMLPGES 4681
QY 1276 ATLSCEVAQAQT-EVTWYKDGKLSSSSKVRTEAGCWRLVVOAGQADAGEYCEA-- 1332
Db 4682 AELHCKLKGSPVQVTFWFKNKELESNTMYFNSEAILDITDKVEDSGSYSCAEVN 4741
QY 1333 --GGQRLSPHLDOVBPKFAKFAKEQLAHRKVOAEAGAIATLSCEVA-QAOTEVTWYKDGK 1389
Db 4742 DVGSDSCSTEIVIKPPSPFIKLEPA---DIVRGTNALLQCEVSGTGPFPISEFKDKKQ 4797
QY 1390 LSSSKVRNEAVGCTRRLLVVOAQCOADTGEYSC---EAGGORLSFLDVAEPKVVFAKE 1445
Db 4798 IRSSKRYLFSQKSLVLEIFSNAGDYGEYCVVANEVKGCGCMATHLLKEPTEFVKV 4857
QY 1446 QPVHEVOQAQASTLSCEVAQAQ-TEVMWYK-----DGK-KLSFSSKVRNEAVGCT 1496
Db 4858 D-----DLIALGQVTLQANVSGSEPIVWTKMGQEVIREDGKIMFSNGVAV----- 4907
QY 1497 RRLVVQQAQADAGEYSC---EAGSQRSLSHLHVAEPKFAKFAKEQPASREVOEAGTSA 1552
Db 4908 --LIIPDVQISFGKYVCLAEENAGSQTSGELIVKEPAKIER---AELIQVTAGDPA 4961
QY 1553 TILSCVA-QAOTEVTWYKDGKLSSSSKVRNEAVGCTRRLLVVOEAGQADAGEY-----SCK 1607
Db 4962 TLEYTVAGTELPKWKYKDGRLVASKYRISFKNNVAQLFYSAELHDSQYTFEISNE 5021
QY 1608 AGDQRLSHLHVAEPKFAKFAKEQPAHREVOEAGASATLSCEVA-QAOTEVTWYKDGK 1666
Db 5022 VGSSSCETFTVLDRIAPFTFKPL-RNVDSVVGTCRLDCKIAGSLPMRVSWFKDGKEI 5080
QY 1667 SSSSKVRNEAVGCTRRLLVVOQAQADAGEYSCA---GGQRLSPRLHVAELEPOISERP 1722
Db 5081 AASDRYIAFVEGTASLEIRVDNMNDAGNFTCRATNSVGSKDSSGALIVQS-PFSFVTKP 5139
QY 1723 CREPLVVKHEBDIILTATPASAATVTLKDGVEIRRSKRHETASQGDTHTLVHGAQ 1782
Db 5140 GSKD---VLPGSAVCLKSTFGQSTPLTRNFKNGKELVSGSCYITKEALESSELYLVK 5196
QY 1783 VLDSAIYSCRVG--AEGQDPVQ--VEEVAAKFCRLLEPVCCEL---GGTVTLACELS-P 1834
Db 5197 TSDSGTYTCKSVNAGGVECSANLFSKE-PATFVEKLEP--SQLLKGDATQLACKVTGT 5253
QY 1835 ACAEVVRCGNTQPVGRKFM--VAEGPVRSLTVJGLRAEDAGEYCESRD---DHTS 1888
Db 5254 PIKITWANDREIKESSKHRMSFVESTAVILRLTDVGI--EDSGEYMCQAEAGSDHCS 5311
QY 1889 AQLTV-SVPRVVKFMSGLST-----VVAEGGEATFCQVVSFSDVAVVMFRGALIQP 1940
Db 5312 SIVIVKESPYTFKEFKPIEVLKEYDVMLLAEVAGTTPFE-----ITWFKONTILRS 5362
QY 1941 SKFALISOGASHSTISDLVLEDAQO-----ITVEABGASSAALVRNEAPVLFKKLEP 1996
Db 5363 GRKYFTFIQDLHLVSQILKFVAADAGEYQCRVTNEVGSSICSAVTLREPP-SPIKKIES 5421
QY 1997 QVVEERSVTVLELTPWEP-ELRWTRNATALAPCKNVEIHAEAGARHLVLHNVGFADRG 2055

Db 5422 TSSLRGCGTAAQATKGLSLPITVTLWKDSDTEITDDNIRMFTFENNVASLYLSGIEVKHGD 5481
QY 2056 PFGCETEDD---KTOAKLTVEMRQVRLVRLGLOAVEREOGTATMEVOLSHA-DVDSGWT 2110
Db 5482 KYVCOAKNDAGIORCSALLSVK-EPATITEEANSIDVTQODPATLQVFGSGTKEITAKWP 5540
QY 2111 RDGLRFOQGPCHLAVRGPMTLTLTSLGRPEDEGLMVFKAEG---VHTSARLVVTEL-- 2164
Db 5541 KDGQELTLGSKYISVTDVTSILKIISTEKKDSGEVTFEVQNDVGRSSCKARINVLDLII 5600
QY 2165 PVFSRPLQDVVITEKEKVTLECEL--SRPNVDVRLKDGVELRACKTMAIAQAQACRSL 2222
Db 5601 PPSFTKLLKNDSTIKGSFIDLECIAGSHP-LSIQMPKDDQDEISASEKEYKFSFHDNTAFL 5659
QY 2223 TIYRCEPADQGVVYCAHDAQSSASVKVQGRTYTLIYRRLAEDAGEIQFAENASRAQ 2282
Db 5660 EISLEGTDGSGTYCSATNK-----ACHNQCSGH 5688
QY 2283 LRVKELPVTILVRPLRDKIANEKIRGULECOVSRASQVRFKFSQBLQPGPKVELVSDGL 2342
Db 5689 LTVKEPPYFVEKQSQSDPNENTRVLKALVGGTAPMTIKWFKONKELSHLSGAARSVMKDDT 5748
QY 2343 YRKLIIISDVHAEDEDTVTC---DAGDVKTSQAFVVEE-----QSTIVVGLQDVTV 2390
Db 5749 STLSLELPAKATDSGTIYICQLSNDVGTATSKATLVKEPPQPIKXPSVLVLNGQSTT- 5807
QY 2391 MEPAPAWFECE-TSIPSVRPPKMLGK--TVLOAGG-----NVGLEBQGTVH 2434
Db 5808 -----FECQITGTPKIRVSWYLDGNEITAIKQHGISFIDGLATFQISGARVENSCTY- 5859
QY 2435 RLMLRRCTMTGPVHFTVCKSRSSAELVSDIPVVLTRPLEPKTGRELSVVLSCDFRP 2494
Db 5860 -----VCEARN-----DAGTASCSIELKVEPPTI-RELKPEVVVKYSDVLECEVTG 5907
QY 2495 APK-AVOWYKDDTFLSPSEKFKMSLEGQMAELRLMLPADAGVYRC-----QAGSAHSST 2549
Db 5908 TPFPEVTWLNKXNREIRSKKYTLTDRVSVENLHITKCDPSDTGEYQOCIVNNEGSCSCT 5967
QY 2550 EVTVE-----AREVTVGPLODAEATEEGWASFCBSLHSHEDBEVHSLNGMPLYNFSFHE 2604
Db 5968 RVALKPSPFIKIENTTTVLKSSATFQSTVAGSPPIS-----ITWKDQDILDEDDNVY 6022
QY 2605 ISHKGRRHTLVLKSIQRADAGIVRASSLKYSTSARLEVRV-----KPVVFLKALD 2654
Db 6023 ISFVDSVATLIQIRSVNDHSGH-----RYTQAKNESGVERCYAFLLVQEPQAVIVEKAK 6075
QY 2655 DLBAERGTLALOCYS-DPEAHVVRKQVGLQPSDKYDFLHTAGTGRGLVHVDVSPEDA 2713
Db 6076 SVDVTEKDPMTLECVVAGTPELKVWLKDGKQIVPSRYFMSFENNVAIFRQSVNMKQDS 6135
QY 2714 GLTYCHV---GSEETRARVRVHDLHV--GITKRLKTMVELEGESCFECVLSHESASDP 2767
Db 6136 GQYTFKVENDFGSSCDAYLRVLDQNIPIPSFTKLTMDKVLGSSIHMECKYS-GSLPIS 6194
QY 2768 AMWTVGKTGVGSSSRFCATQGRKYILVVREAAPSADAGEVVFVSRGLTS-----KASLIYR 2823
Db 6195 AQWFKDQKEISTSAKRYLVCHERSVSLEVNLEEDTANYTCKVSNVAGDADACSGILTIVK 6254
QY 2824 ERPAALIKPLEDOWAPGEDVELRCELSRAGTP---VHMLKDKKARKKQKQVVDVCEGTM 2880
Db 6255 EPSFVLVKGROQAI-PDSTVEPKALK--GTPPFKIRKWKDQDELVSFGPKFCIGLEGST 6311
QY 2881 AMLVIRGASLIKDAGETYC---EVEASKSTASHVEEKANCECTEELTNLQVEEKG-TAYF 2935
Db 6312 SPLNLYSDASKGTQYTCVTNDVGDSDCTTMLLVTEPPK-FVKKLEASKIVKAGDSSRL 6370
QY 2936 TCK-THEPATYTVWRKGLLELRASGHQPSQEGTLRLITISALEKADSDTYTCD-----IG 2990
Db 6371 ECKIAGSPEIRVWVFNHEHLPASDKRYMTFIDSVAVIQMNNLSTEDSGDFTCEAQNPAQ 6430
QY 2991 QAQSRQALVQGRRVH-----IIEDLEDVDVQEGSGSATFRCRISPANYPEVHFWFLDKTPL 3045
Db 6431 STCSCTKIVKPEPVSFSSPPPIVETLKNAEV-----SLECELSGTPPEVWYWKDQRL 6484

| | | |
|------|---|------|
| 8482 | VBPNAGKYTCQIKNDAGWQECFATLSVLBPATIVEK-----PESi | 8522 |
| 5117 | PSLLTSDAAPVFLTELQNEQVQDGYPSFDCVVTGQMPVSRVWFKDGKLLBEDDHMYIN- | 5175 |
| 8523 | -----KVTTGDTCTLECTVAGTPELSTKFWKDGKELTSDNKYKISF | 8563 |
| 5176 | EDQCGGHOLIITAVVPADMGVYRCIAENSGVSTKAELEAVDLTSTDYTAADATBESSY | 5235 |
| 8564 | FNKVSQ--LKLIINVAPSDGVSFEVQNPVKDSCITASQV----- | 8602 |
| 5236 | PSAQYLSRSRBEQGETSTDEGQLPQVVEBLRDQVAPGTRLAKFQKVKGYAPAPRLYWF | 5295 |
| 8603 | -----SDRTVPSPFTRKLKETNGLSGSVV-MECKVYGPPIISGVWF | 8643 |
| 5296 | KDQGLPTASAHIRMTGKKILHTLEIISVTRSDSOYAAIYSNAGAAYSABLLVGRPDE | 5355 |
| 8644 | HEGNEISSGRKYQTLTNTCALTVNMLESDSGDYCIATNAGSDECSAPLTVR---- | 8699 |
| 5356 | PEEKPADSVHQBVLPPRLMEFTFKYKKGSSITFSVKVEGRVPVTVHMLREABERGVLW | 5415 |
| 8700 | -----EPPSFQKPDMDVLGTGNTVTF--SIVKGTTPPFSVWFKGSSEL---- | 8742 |
| 5416 | IGPDTPG--YTVASAAQSHSLVLDVGRQHQGTTCIASNAQAOLCSASLHVGLPKVE | 5473 |
| 8743 | ----VPGDRCNVLSDESVAELEFDVDTOSGEYTCIVSNEAGKASCITHLIKAPAKFV | 8798 |
| 5474 | EQ-----EKVKEALISTFLOGT--TOAISA-----QGLETASFADLGGORKEEPLAKE | 5520 |
| 8799 | KELANDYSIEKGPIL-----LEGTTGTPPISVTKKNGINVP-----SORCNIITTEKS | 8849 |
| 5521 | ALGHLSLAEVQTEE-----FLOKLTQSITEMWSAKITQAK-----LQVPGGSD | 5564 |
| 8850 | AI--LEIPSTVEDAGQVNCYIENASGK--DSCSAQILILEPPYFKVLEPVKVSVDGA | 8905 |
| 5565 | EDSKTPSAPRHGRS-----RPSSS-----IQES-----SSRSE | 5593 |
| 8906 | SLQQLAGTPIGVSWKYGTDKTRPPTTYTKMFFRNNVATLVFNQVDINDSGYICKAENS | 8965 |
| 5594 | DGDARGEIDFYVVTADYPLGABQDAITLREGQYVEV-----LDAAHPL--RWLVTRKP | 5646 |
| 8966 | VGESASATF--LTQEQKLPSPFSRQLRDVQETVGLPVVFDCAISGESEIYSWYKDGKP | 9023 |
| 5647 | TKSPSRGCVSPAYLDRRLKLSPEWGAABAPFPGEAIVSEYKARLSVTOELLSEQ | 5706 |
| 9024 | LKDSFN-----VQTSFLDNTATLN-----IFKTRSLAGQYSCYATNPGSSASGAR | 9070 |
| 5707 | AFVBELOQLQSHLQHLERCPHVPIAVAGKAVI-----PRN | 5743 |
| 9071 | LILTE-----GKNPPFFDIRLAPVDVAVGVSADPECHVTGTQIKVSNWAKDSRE | 9119 |
| 5744 | VRDIGRHSFLOE-----LQCDTDDVAMCFIKNQAAPEQYLEFLVGRVQAESVYVST | 5798 |
| 9120 | IRSGKQIYSILENSAHLTVLKVDBKGS-----GQYTCVAVNEVGKDSCTAQL | 9167 |
| 5799 | AIQRFYKYAEALLAGDPQPPPPLOHYLEQVPERVQRYQALLKELRNKARNQNC | 5858 |
| 9168 | NIKE-----KLIPPSSTFKLSETVEET----- | 9189 |
| 5859 | LLEQAYAVSALPORAENKHLVSLMENYPTLEALGEPIRQGHFIWVGAPGARM----- | 5914 |
| 9190 | -----EGNSFKLEGRVAGSQPIITA | 9209 |
| 5915 | WGHNRHV-----FLFRNLVTCRPRDSRTDTSYVFRNMKLLSSDLNDQVEGDGR | 5967 |
| 9210 | WYKNNIEQTSNCEITFKNNTLVQRKAGMDAGLYT-----CKVS-----ND----- | 9254 |
| 5968 | AFEVQEREDSVRYLLOARTAILKSGW--KEICGIQOORLALPVNRPPDFEELADCTAE | 6026 |
| 9255 | -----AGSNLCTSSIVKE-----PKAPPVFDHLLTPVTVS | 9285 |
| 6027 | LGETVKLACRVGTGPKPVISWKDGKAVQDPHHLLIEDPDGSCALILDSLTGV----- | 6080 |

QY 6887 -----LARGSRHPPARRHLLKGGYIAGALPGLREPLMEHRV-----LE 6925
 DB 10328 EIPKKVPEERKPPVKEEYPPPPKVPALP--KPVPEEKVAVVPVAKAPPRAEVS 10385
 QY 6926 BEAAREOATLAKAPSETALRLPASGTHLAPGH--SHSLEHDSPS-----6970
 DB 10386 KKTVEEKRFVAEEKLSFAVQRVETVREHSABEENSYSSEEGVLSIVREBEREE 10445
 QY 6971 -----TRPSSEACGE-----AQLPSAPSGGAPIRDMGHPPQSK-----QLPSTG 7011
 DB 10446 EAEVTEYEMBEPEYVVEEKHLHISKRVEAEP-----EVTROEKKIVLKPKPAKI 10499
 QY 7012 GHGTAQ--PERSPDSNGO--PARCPKOGSAPQCGSPHAPVAPCPGSPFPFGCK-E 7068
 DB 10500 BEPPPAKUPEAPKTVPEKKVAPV--PKKEKVPKVPPEPK-KPVPEKKVPPKVKME 10556
 QY 7069 APL-----VPSSPFFZGQOAPAPAPAXASPPPLDSKMG--PDISLPGRPKPGPCSS 7116
 DB 10557 EPLPAKVTERRHQITQEEKVLVAVTKKAPPKARVPEPKAVPEEKVYLKPKKREB-EP 10615
 QY 7117 PCSASQASSQVSSLRVSSQVGTGTPGSLDAEGTQEB--AEDLSDTPTLORPQOATM 7174
 DB 10616 PAKVTEFRKRVVKEEKVSIAPKREFPQIKVTIMEEKERAYTLEEBAVSVQREEEY 10675
 QY 7175 RKFSLGGRGGVAGVAGYGTFAFGDAGGMLGQGPMMARIAVAVSQSEEEQEEARARSQS 7234
 DB 10676 EYDYKEEYEPTEEDYQY-----EYEEEREYERYEHE 10710
 QY 7235 EQQARAEPLQVQASRPVVEVGRAP-----TRSSPE---PTWEDIGQVSLVQIRDLS 7286
 DB 10711 EYITEPEKPIPKVPVEBPVPTKPKAPPAKVLKKAAPVEKVPV-----10754
 QY 7287 GDAAEADTISLDSVDPAYNLSLDYDKYLPFEFMIFRKPVKPSAOPPEPPSPMAEBELA 7346
 DB 10755 -----IPKLLKPPPK-VPEBPK 10772
 QY 7347 EPEPTWMPGELPHAGLEITE-BSEDV-----DALLAAAVGRKRKWSPSR 7394
 DB 10773 VPEE-----KIRISITKEQVTEPAAKVPMKPRVVAEEKVPVPRKEVAP-- 10819
 QY 7395 SLFHPGR--HLPLD-EPAELGLRERVKASVEHISRII-----7429
 DB 10820 -----FVRVPEVPEKELEPEEVAEEV---VTHVEEYLVEEVEEYIHEEEFITEBEVVP 10871
 QY 7430 -----KGRPEGLEKEGPPPKKP 7446
 DB 10872 VIPKVPVEPRKVPVEKPKVPVPPKKEAPKVP 10906

RESULT 10
 Q8NH4 PRELIMINARY; PRT; 658 AA.
 ID Q8NH4;
 AC Q8NH4;
 DT 01-OCT-2002 (TReMBLrel. 22, Created)
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Obscurin (Fragment).
 GN OBSCN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Young P., Ehler E., Gautel M.;
 RT "Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere
 RT assembly".
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ314904; CAC85751.1; --
 DR InterPro; IPR003599; IG--
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003598; IG_C2.

DR Pfam; PF00047; ig; 4.
 DR SMART; SM00409; IG; 6.
 DR SMART; SM00406; IGC2; 3.
 DR PROSITE; PS0835; IG_LIKE; 4.
 KW Immunoglobulin domain.
 FT NON_TER 1
 FT NON_TER 658 658
 SQ SEQUENCE 658 AA; 72830 MW; CB37B453FCD4AE4A CRC64;
 Query Match 8.2%; Score 3389; DB 4; Length 658;
 Best Local Similarity 100.0%; Pred. No. 1.2e-173;
 Matches 658; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2075 ROVLVRGLQAVAEAREEQGTATMEVQLSHADVDGSTRDGLRFQOQPTCHLAVRGPMTLT 2134
 DB 1 RQVLVRGLQAVAEAREEQGTATMEVQLSHADVDGSTRDGLRFQOQPTCHLAVRGPMTLT 60
 QY 2135 LSLRPEDSGLMVFAEGVHTSARLVVITELPVSFSRPLQDVVTEKEKVTLECELSRNV 2194
 DB 61 LSLRPEDSGLMVFAEGVHTSARLVVITELPVSFSRPLQDVVTEKEKVTLECELSRNV 120
 QY 2195 DYRWLKDGVLELRAGKTMAIAAQACRSITIRCFEADQGVVCDADHAQSSASVKVQGT 2254
 DB 121 DYRWLKDGVLELRAGKTMAIAAQACRSITIRCFEADQGVVCDADHAQSSASVKVQGT 180
 QY 2255 YTLIYRRLVLAEDAGBIQFVAENAESRAQLRVKELPVLVRLRDKIAMKRGVLECOVS 2314
 DB 181 YTLIYRRLVLAEDAGBIQFVAENAESRAQLRVKELPVLVRLRDKIAMKRGVLECOVS 240
 QY 2315 RASAOVRWFKGSOELQPGPKYELVSDGLYRKLIISDVHAEDEDVTCDAQDVKTSAQFFV 2374
 DB 241 RASAOVRWFKGSOELQPGPKYELVSDGLYRKLIISDVHAEDEDVTCDAQDVKTSAQFFV 300
 QY 2375 EQSITIVRGLQDVVMEPAPAFWECETSI PSVRPPKLLGKTVLQAGNVGLEQGT VH 2434
 DB 301 EQSITIVRGLQDVVMEPAPAFWECETSI PSVRPPKLLGKTVLQAGNVGLEQGT VH 360
 QY 2435 RLMRLRTCTMTGPHFTVTKSRSSARLVSDIPVLTPLPTEPKTGRELQSVLSCDERP 2494
 DB 361 RLMRLRTCTMTGPHFTVTKSRSSARLVSDIPVLTPLPTEPKTGRELQSVLSCDERP 420
 QY 2495 APKAVQWYKDDTFLSPSEKFKMSLEGQMAELRLIMPADAGVYRCQAGSAHSSTVTE 2554
 DB 421 APKAVQWYKDDTFLSPSEKFKMSLEGQMAELRLIMPADAGVYRCQAGSAHSSTVTE 480
 QY 2555 AREVTVTGLODAEATEEGWASFCELSHEDDEWWSLNGMPLYNDSFHEISHKGRHTL 2614
 DB 481 AREVTVTGLODAEATEEGWASFCELSHEDDEWWSLNGMPLYNDSFHEISHKGRHTL 540
 QY 2615 VLKSIQRADAGIVRASSLKVSTASRLVVRKVPVFLKALDLSASERGTALQCEVSDPE 2674
 DB 541 VLKSIQRADAGIVRASSLKVSTASRLVVRKVPVFLKALDLSASERGTALQCEVSDPE 600
 QY 2675 AHVVRKDGVLGSPDKYDFLHTAGTRGLVHVDVSPDAGLYTCHVGSEETRARVRVH 2732
 DB 601 AHVVRKDGVLGSPDKYDFLHTAGTRGLVHVDVSPDAGLYTCHVGSEETRARVRVH 658
 RESULT 11
 ID Q10465 PRELIMINARY; PRT; 7962 AA.
 AC Q10465;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Elastic titin (Fragment).
 GN TITIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

| | | |
|----|--|--|
| RC | TISSUE=Skeletal muscle; | |
| RX | MEDLINE=96026330; PubMed=7569978; | |
| RA | Labelit S.; Kolmerer B.; | |
| RT | "Titins, giant proteins in charge of muscle ultrastructure and | |
| RT | elasticity."; | |
| RL | Science 270:293-296(1995). | |
| DR | EMBL; X90569; CAA62189.1; - | |
| DR | PIR; I38346; I38346. | |
| DR | GO; GO:0030017; Cisarconere; TAS. | |
| DR | GO; GO:0008307; Pistructural constituent of muscle; TAS. | |
| DR | GO; GO:0006941; Pistructured muscle contraction; TAS. | |
| DR | InterPro; IPR007110; IG-like. | |
| DR | InterPro; IPR003598; IG_c2. | |
| DR | InterPro; IPR004168; PPAK_motif. | |
| DR | Pfam; PF00047; ig; 59. | |
| DR | Pfam; PF02818; PPAK; 53. | |
| DR | SMART; SM00408; IGC2; 43. | |
| DR | PROSITE; PS50835; IG_LIKE; 58. | |
| FT | NON_TER 1 | |
| FT | NON_TER 7962 7962 | |
| SQ | SEQUENCE 7962 AA; 883018 MW; B85240533CBAD58 CRC64; | |
| | Query Match 8.0%; Score 3313; DB 4; Length 7962; | |
| | Best Local Similarity 21.5%; Pred. No. 1e-167; | |
| | Matches 1710; Conservative 1105; Mismatches 3287; Indels 1868; Gaps 306; | |
| QY | 10 PRFLTRPKAFVSVGKDATLSQIVGNPTPPQVSWKXQOPVTAGARFRLAQDGLYRLTI 69 | |
| DB | 10 PALIPLQDTVSEGPARFQCRVSGTDL-KVSWYKOKKIKRPSFRFTQFEDTYQLEI 68 | |
| QY | 70 LDALGDSQYVCRAANAIGERAFPAVGLQVDAEACAEQ-----APHLLRPT 117 | |
| DB | 69 AEAYPEDEGTYTFVANNVAGVQSVSTANLSLEAPESILHERIBQEIEMEMKAPVIRKIE 128 | |
| QY | 118 SIRVRGSGATRCRGGSPRAVSGSKDGRGLGPDGPRVVEBELGASALRIARAAPR 177 | |
| DB | 129 PLEVALGHIAKTCETCIOSAPNRFQWFKAGRIEYSDKCSIRSKY--ISSLEILLTQV 186 | |
| QY | 178 DGGTYEVRAENPLGAASAAAALVDSDAADTASRFGTSTAALLAHQRRREARAGAPA 237 | |
| DB | 187 DGEYTCASNEYSVSCATLTV-----TEA 213 | |
| QY | 238 SPSTCTRTCTVTE--GKARLSVYTGEPKPTVMKDGQLVTEGRRHVYEDAEQNFV 295 | |
| DB | 214 YPTFTLSRPKSLTFFVGAAKTCTVTGT-PVITETWQDGAALSPSPNWI--SDAENKHI 272 | |
| QY | 296 LKILFCQSDRGLYTCTASNLVG-QTYSSVLVVRPAPVFPKRLQDLREVREKESATFLC 354 | |
| DB | 273 LELSNIQIDRGVYSCASKNFGADICQAEIILIDKP--HFKELEPVQSAINKVHLEC 330 | |
| QY | 355 EYDQP-STENAMPKEETRLWASAKYIEEGTERLTVRNVSADDDAVYICETPE---GS 410 | |
| DB | 331 QVDEDRKVTWNSKQGLKPPGKDYKICFEDKIATLEIPLAKLQSGTYVCTASNEAGSS 390 | |
| QY | 411 RTVAELAVQ--GNLLRKLPRKTRVRVGDTA-MFCVELAVPGVPVHWRNQBWEVAGGRA 467 | |
| DB | 391 SCSTVTVREPPSFVKVPDPSYMLPGEARLHCKLKGSPVQVTFWKNKELSESNTVR 450 | |
| QY | 468 ISAEGRHTLITISQCCLEVGQVAFWA----GDCQTSRTRFCSAPKPKPLQPPVPVKA 523 | |
| DB | 451 MYFVNSEAILDIDTVKVEDSGSYSCAVANDVSDSCSTEIVIKEP--PSFIKTLFPAFIV 508 | |
| QY | 524 RMESVILSWSPPHGERPVTIDGVLVEKKLGLTYTWIRCHEAEMVATELTVADVABEG 583 | |
| DB | 509 R-----GTNALLQC-----EVSGTG 523 | |
| QY | 584 NFQFRVSALNSFQSPYLEPFGTWHLAPKLAVRTPLKAVQAVEGGEVTFSDLTVASAGE 643 | |
| DB | 524 PFEI-----S 528 | |
| QY | 644 WFLDGOALKASSVYEHCDTRHTLPIREVPSALHGAQLKFVANGTIESIRVEVRAAPGL 703 | |
| DB | 529 WPKDKQIRSSKKYRLFQSKLVCLEIFSFNSADVGSEYSCVWANEVKGCGCWATH----- 583 | |

| | | |
|----|---|--|
| QY | 704 TANKPPAAAAAR--EVLARLHHEEAQLLAEL--SDQAAAATWLKDGRTLSPGPKYEVQASAGR 760 | |
| DB | 584 LLKEPPTFFVKVVDLIALGCGTQVTLQAAVRGSGSPISVTWKGQEVIREDKIKIMSSNGV 643 | |
| QY | 761 RVLVRDVARDDAGLXECVSR---GGRIAY-QLSVQGLARFLHKDMAGSCVDVAVAGPAQ 816 | |
| DB | 644 AVLIIPDVQISFGGKYTCLAENEAGSOTSVCGLIVKEPAKIER--AELIQVTAGDPAT 700 | |
| QY | 817 FEETS-EAHVHVHVKYDGMELGHSGERFLOEDVGTGTHRLVAATVTRQDRGTV----SCR 871 | |
| DB | 701 LEYTVAGTPELKPWKYKDGRLPLVAS-KKYRISPKNNVAQLKFYSAEHLHDSGQTFEISNE 759 | |
| QY | 872 VGEDSVDFRLRVSEPKVV--FAKEQLARRKLOAEAGASATLSCEVA-QAOTEVTVYKDGK 928 | |
| DB | 760 VGSSSCETTTFLDRDIAFFTK--PLRNVDVSVNGTCRLDCKIAGSLPMRVSWPKDGK 816 | |
| QY | 929 KLSSSKVCWEATGCTRLRVVQOAGADAGEYCEA-----GGORLSPHLDVKEPKVPV-- 982 | |
| DB | 817 EIAASDRYRIAFVEGTASLEIIRVDMNDAGNFTCRATNSVGSKDSGALLVQEPFSPVTK 876 | |
| QY | 983 --AKDOVAHSEVQAEAGANATLSCEVAQAQAE---VMVYKDGKKLSSSLKRVHAEAGCR 1036 | |
| DB | 877 PGSKDVLPGSVAV-----CLKSTFGSTPLTRWFKGNKELVSGSGCVITKEALE 925 | |
| QY | 1037 RRLVVQOAGKTADGYSCE---ARGQVSRFLHITTEPKMFAKEQSVHNEVQAEAGASA 1092 | |
| DB | 926 SSLELYLVKTSDSCTYTCVSNVAGVECSANLFVKEPATFVEK---LEPSQLLKGDAT 982 | |
| QY | 1093 MLCSEV-AQAQTEVTVYKDGKLSSSKVGMEVKGCTRLRLVLPQAGKADAGEYSCAAGG 1151 | |
| DB | 983 QLACKTGTTPPKITWPFANDREIKESKHKRMSFVESTAVLRDLTDVGLTDSGEYCEAONE 1042 | |
| QY | 1152 RVSEPH---LHITPEKGVFAKEQSVHNEVQAEAGTAMLSCEVA-QPOTEVTVYKDGKKLS 1207 | |
| DB | 1043 AGSDHCSSIVIVKESPYFTEK---FKPIELVKEDVNMLLAEVAGTPEFTEWFKDNTILR 1099 | |
| QY | 1208 SSKVRNEVKGCTRLRVVQVQKADAGEYSC---BAGQVRSFQLHITTEPKAVFAKEQL 1263 | |
| DB | 1100 SGREYKTFIQDHLVSLQILKFPVADAGEYOQVTVNEVSGSICSAVTVLTPRPPSPFKK--- 1156 | |
| QY | 1264 VNEVTRTEAGASATLSCEVAQAQTEVTVYKDGKLSSSSKVRIEAAQCMQLVQQQAQA 1323 | |
| DB | 1157 IESTSSLRGGTAFOATLKGSLPITVTLWLDKSDIEDTDNIRMTFENNVAASLYLSGIEVK 1216 | |
| QY | 1324 DAGETC-----EAGQRLSFHLVSEPKAVFAKEQLAHRKQVQAEAGATLSCEVAQAQT 1379 | |
| DB | 1217 HDGKYVQQAQNDAGIQRCSSALLSVKEP-ATITBEAVS---IDVTQGDPAQLQVKSFTK- 1271 | |
| QY | 1380 EVT--WKDGKLSSSSKVMEAVGCTRLRVVQQAQCAQDTEY-----SCRAGG 1426 | |
| DB | 1272 EITAKWFDQOELTGLSKYKISVTDVTSILKIISTEKDSDGEYTFEVQNDVGRSSCKA-- 1329 | |
| QY | 1427 QRLSFLDLVABPKVYFAKEQPVHREVOQAQAGATTLSCSEVAQAQ-TEVMYKDGKKLSFS 1485 | |
| DB | 1330 -RINV-LDLIIPP-SFTKKL---KMDSIKSGSFIDLECIVAGSHPIQWFKDQDEISAS 1383 | |
| QY | 1486 SKVMEAVGCTRLRVVQQAQADAGEYSC---EAGQRLSFHLHVAEPKAVFAKEQOPAS 1541 | |
| DB | 1384 EKYFSPHDNTAFLEISOLEGSDGTGTCATNAGHNQCSGLTVKEPP--YFVEKQPS 1441 | |
| QY | 1542 REVQAEAGTSATLSCEV-AQAQTEVTVYKDGKLSSSSKVMEAVGCTRLRVVQEAQAD 1600 | |
| DB | 1442 QDV--NPNTRVQLKALVGGTAPMTIKWFKDKNELHSGAARSVMKDDTSTSELPFAAKATD 1499 | |
| QY | 1601 AGEYSCK-----AGQRLSFHLHVAEPKVPFAKQPAHREVOQAAGASATLSCEV-AQAQT 1655 | |
| DB | 1500 SGTYICQLSNDVGTATSKATLFFVEKPPQFTFKKPSV---LVLNQGQSTTFECQITGTPKI 1556 | |
| QY | 1656 EVTVYKDGKLSSSSKVMEAVGCTRLRVVQQAQADAGEYSC---AGQRLSFHLH 1711 | |
| DB | 1557 RVSNYLDGNEITAIQKGIISFIDGLATFQISGARVENSIGTVVCEARNDAGTASCSELVK 1616 | |

QY 1712 AELEPOISERPCRRBPLVYKEHEDILATLATPSAATVWMLKOGVEIRRRKQKHETASQG 1771
 DB 1617 KEPPFTFIRE---LKPVEVVKYSDVELECEVTGTPPFEVTLWKNREIRSKKXKTLTDRV 1672
 QY 1772 DTHLTIVHGAQVLDGAIYSCRYGASGODFPVQVEVAAK---FCRLLEPVCGLGGTVT 1827
 DB 1673 SVFNLIHTKCDSDTGEYCIVSNEGSCST-RVALKEPSPFIKKIENITTVLKSSAT 1731
 QY 1828 LACEL---SPACAEVWRCNTOPRVGKQFQWVAEGP-----VRLTIVLGLRAED---AG 1876
 DB 1732 FOSTVAGSPPIIS-ITWL-----KDDQILDEDDNVVISFVDSVATLQIRSVDNHSG 1781
 QY 1877 EYVCESRD---DHTSAQLTVSVPRVVFMSGLSVVAEEGEATFQCVWSPS-DVAVW 1931
 DB 1782 RITCQAKNESGVERCYAFLLVQEP--AQIVEKAKSVDTVEKDPMTLECVVAGTPELVKVM 1839
 QY 1932 FRDGALLOPSEKFAISQSGASHSLTISDLVLEDAQITVEAE---GASS-SAAALRV--RE 1985
 DB 1840 LKDGKQIVPSRYFSMFENNVAFPRIQSVMKQDSQYTFKVENDFGSSSCDAYLRVLQDN 1899
 QY 1986 APVLEFKKLEPOTVEERSVTLVLELTPWP-ELRWENATALAPGKQVEIHAEGARHL 2044
 DB 1900 IPPSFTKLTKMDVKVLSIHCKVSGSLPISAQWFXD-----GKEI---STSAYRL 1950
 QY 2045 VLNWVGADRGFGCETPDKTQAKLTVMEMQVRLVRLQAVEARECQTATMEVQLSHAD 2104
 DB 1951 VCH-----ERSVSEVNLEL-----EDTAN----- 1971
 QY 2105 VDGSWTRDGLRQCOQPTCHAVRGPWHLTJLSGLRPEDSGLMVFAEGVHTSABLVTTEL 2164
 DB 1972 -----YTKVSNVAGDDA-----CSGILTVEKP 1994
 QY 2165 PVSFRP-LQDVV--TTEKEKVTLECELSPRNVVRLKDGVELPAGKTMATAIQAQGRS 2221
 DB 1995 PSFLYKPGQQAIPQSTVEFKALK---GTPPFKIKWFKDDVELVSGPKFCIGEGTSF 2051
 QY 2222 LTYICEFADQGVYCDADHAQSSASVKVQGRYTLIYRRVLAEDAGBIQFVAENASRA 2281
 DB 2052 LNLYSVDASKTQYCTCH-----VTNDVG-----SDSCCT 2080
 QY 2282 QLRVKELPVLVRLPRLDKIAMEKHGVLCOVSRASAQVR--WFKGSOELQPGKVLVS 2339
 DB 2081 MLLVTEPPKFKLEASKILVKAQSSRLCEKIA-GSPEIRVWVFNHEHLPASDKYRT- 2138
 QY 2340 DGLYAKLIISDVHAEDEDTYTCADGDVKTSAQFFVEEQSITIVRGLQDVTVMPEPAPWE 2399
 DB 2139 -----FIDSVAVIQWNNLSTEDSGD-----FI 2160
 QY 2400 CETSIPSVRPPKVLIGKTVLQAGNVGLQEGBTVHRLMLRTRCSTMTGPHFTVGKSRSS 2459
 DB 2161 CEAQNP-----GST-----SCST----- 2174
 QY 2460 ARLVSDIPVVLTRPLEPKTGRELOSVVLSCDFRPAPK-AVQWYKDDTPLSPSEKPKSL 2518
 DB 2175 -KVIKEPVPFSFPPIVETLKNAE-VSLECELSCTPPFEVVWYKDKQLRSSKKYKIAS 2232
 QY 2519 EQOMAEILRLIMPADAGYVRCQA-GSAHSSTEV-TVEARE-----VTVTGPLQ 2565
 DB 2233 KNFHTSIHLNVDTSIDIEYHCKAQNEVGSCTCVTKLEPPRPFSVKLSLTVVAGEPA 2292
 QY 2566 DAEATEEG-----WASFCELSHEDVEWMSLNGMPLYNDSFHEIHKGRHTLVLS 2618
 DB 2293 ELQASIEGAQPIFVQWLEKEVIRESENI-----RITFVENVATLOFAK 2337
 QY 2619 IQRADAG-----IVRASSLKVSTSRARLEVRVKPVVFLKALDLSABEGLTALQCEVS-D 2672
 DB 2338 AEPANAGYICQIKNDGGR-ENMATLMV-LEPAVIVEKAGPMTVTGECTLECKVAGT 2395
 QY 2673 PEAHVVRKDGVLGSPDKYDPLHTAGTRGLVVDHVSPEADAGLYC-----HVGSEETAR 2728
 DB 2396 PELSVENYKDGKLLTSSQKHFSFYNKISSRIILSVRQDAGTTFQVQNVVHGSSCTAV 2455
 QY 2729 VRVHDJHV--GITKRLKTMVELEGESCFEVLSHESADPAMWTVGKTVGSSSRFQAT 2796

DB 2456 VDYSRAVPPSFRRUKNTGGVLGASCIILECKVAGSPISVA-WPHEKTKIVSGAKYQTT 2514
 QY 2787 QGRKYLTVREAAPSDAGE---VWFSVRGLTSKASLIVRERPAALIKLELDQWVAPGED 2843
 DB 2515 FSDNVCTQLNSLSDSDMGNYTCTVAANVAGSECRVLTIVQPPSPVKPEPELVLPKN 2574
 QY 2844 VELRCELSRAGTP---VHMLKDKARKQKQYDVCEGTNAMLVIRGASLKADGETVCEV 2900
 DB 2575 VTFTSVI--RGTPPFKVNFRGARELVKGRCNIFYEDITVAELELENIDISQSGEYTCV 2632
 QY 2901 EASKSTAS---LHVEKANCFTTELTNLQVEE-KGTAVFTCKTEHPAATVTRKGLLEL 2955
 DB 2633 SNVAGQASCTTRLFVKEPA-AFLKRLSDHSVEPKSIIILESTYTGTLPISTVWKDGFNI 2691
 QY 2956 RASGRHQPSQEGTLTALTISALEKADSDTYTCDIGOAQRAQLLVQGRVHIIE----- 3009
 DB 2692 TTSEKCNIVTTEKTCILEILNSTKRDAGQVSCBI---ENEAGRDVCGALVSTLEPPYFVT 2748
 QY 3010 DLEDVDVQEGSSATFECRISPANYEPVHWFLDKTLPLHANE-----LNEIDAQPGGYHVL 3063
 DB 2749 ELEPFAAVGDSVSLQCVAGTPEITVSWYKGTUKLRPTPEYRTYFTNNV-----ATL 2801
 QY 3064 TLRQALKDSTGITYFEA---GDORASAAALRVTEK--PSVFSRELTDAITTEGEDITLYC 3117
 DB 2802 VFNKVNDSGEYTKCAENSIGTASSKTVFIQEROLPPSFARQLKDIEQTVGLPVTLC 2861
 QY 3118 E-TSTCDIPMCWTXKDKITLRSARCOLSHEHRAQLLITGATLQDSGRYKCEAG---GAC 3173
 DB 2862 RLNGSAPIQVWYRDGVLLRHENLQTSFVDNVATLKILQTLDSHSGQYSCSASNPLGTA 2921
 QY 3174 SSSIVVHARPVR---FOEALKDLEVEGGAATLRCVLSSVAAPVK--WCYGNVNLRRP 3227
 DB 2922 SSS-ALUTAREPKSPFFDIKPVSIDVIAGESADFECHVTG-AQPMRITWKNKEIRPG 2979
 QY 3228 DKYSLRQEGAMELVVNRNLRPDQSGRYSC-----SFGDQTSATLVTALPAQIFOKLR-N 3282
 DB 2980 GNYTITCVNGTPHLRILKVGKDSGQYTCQATNDVKMDCSAQLSVKE--PPRFVFKLEAS 3038
 QY 3283 KEATEGATATLRCELSKT--APVWRKSGSETLRDGRYCLRODGAMCELOIRGLAWDAA 3340
 DB 3039 KVAKGESIQLECKISGSPKIVKSWFRNDSHESKWNKMSFINSVALLTINEASAEESG 3098
 QY 3341 EYSCVC---GEERTSASLTIRPMPAHFTGRRLRHOESIEGATATLRCELSKAAPVE--WR 3394
 DB 3099 DYICEAHNGVGDASCSTALTIVKAPPV-FTQKSPVGLKAGSDVILQCSISGTPPEVWV 3157
 QY 3395 KGRESLRDGRHSLRQDGAVALCLOGLAVADAGEYSCVC---GEERTSALTIVKALPA 3450
 DB 3158 KDRQVNRNKKTKITSKHFDTNLHNLNLEASDVGEYHCKATNEVGSDDTSCSVKPE-PP 3216
 QY 3451 KFTGELRNEAVEGATAMLCVSKYAPVE--WRKGPENLRDGRYILRQEGTR----- 3502
 DB 3217 RFVKLSLDTSTLIGDAVELRAIVEGFPISVVW-----LXDRGEVIRESENTRISFDN 3270
 QY 3503 -CELOQCGLAMADAGEYLCV---CGERTSATLIRALPARFIEDVKNQAEAREGATVL 3557
 DB 3271 IATLOIGSPASNSGKYICQIKNDAGMRECSAVLATLE-PARIIEKPEPMTVTGPNPFAL 3329
 QY 3558 QCELSAAV--EWRKGSSETLRDGRYSRLRQDQTKCELOIRGLAMADTGEYS----- 3607
 DB 3330 ECVVTGTPELSAKFKDGHLSADSXGHTITFINKVASLKI PCAEMSDKGLSYFVKNVSG 3389
 QY 3608 ---CVCGQERTSAMLTVRALPIKFTGELRNEAREATGATAVLRCELSKKAPVE--HWKHE 3662
 DB 3390 KSNCT-----VSVHSDRIVPSPFIRKLKNVAILGASVLECRVSGSAPISVGFODGN 3444
 QY 3663 TLRDGRHSLRQDGAVALCLOGLVAEDAGEYLCV---MCGKERTSAMLTVRAMPSKIE 3718
 DB 3445 EIVSGPKCSSSENVCTNLNLSLSPDTGTITVCAANVAGSDECSAVLTVOEPPS-FEQ 3503
 QY 3719 GURNEATEGDTATLWCELSKAAP--VWVRKGHETLRDGRHSLRQDGRSRLCLOIRGLAV 3776

3504 TPDSEVLEPMSLITFTSVIRGTPPKVFWKSGRELVPGESCNISLEDFVTELELFEVQP 3563
3777 VDAEYSCVCGQERTSAILTVRAL---PARFIEDVNOAREGATAVLOCELSKAAP--V 3831
3564 LESGDYSCLVTDAGSASCTTHLFVKEPATFKRLADFSVETGSPVILEATYTGPPISV 3623
3832 ENRKGETLUGDGRYSLRGDRFCELOHGLSVADTGEYSCV---CGOERTSALTIVRA 3887
3624 SHIKZEYLLISQERCSITWEXSTLILESTIEDYAQYSCLIENEAQDICEALSVLE 3683
3888 POPVREPLOSLOAEESTPATLOCEL-SEPTATVNVMSKGLLOLQANGRRRPLQOCTAEL 3946
3684 P-PYFTEPLEHVEAVIGEPATLQCKVDGTFEIRISWYKEHTKRLPAPAYKMOFKNVASL 3742
3947 VLQDLQREDTGEVTC---TCGSOATSLTWTVA--APVRFURELHQHVEDEGGTAHLCC 4000
3743 VINKVDHSDVGEYSCADNVGAVASSAVLVNKAARLPPFPARKUDVHETLGFVPAFEC 3802
4001 ELSAR-GASVEWRKGSLOLFPCKAYQMVODGAAELLVRGVEQEDAGDYTCDTGHTQMA 4059
3803 RINGSPLQVSWYKDGVLKDDANLQTSFVHNVALTLQILQDQSHIGVNCASNPLGTA 3862
4060 SLVSR-----VPRPKFTRLOSLEOFTDIARLCCQLSDAESGANVQWLKEGVELHAG 4112
3863 SSSAKLILSEHEVP-PFFDLKPVSDVLDALGESGTFKCHVT-GTAPIKITWAKDNREIRPG 3920
4113 PKYEMRSQATRELLIHOLEAKDTGEYAC---VTGGQKTAASLRVTEPEVTVIRGLVDA 4168
3921 GNYKMTLVENTATLTVLVKVGKDGAGQYTCYASNIAGKDCSQAOLGVQBPPEFIKLEPSR 3980
4169 EYVADVEDBFSCEVSRAGATGVQWCLQGLPLQSGNEVTEVAVRDGRHILRLKGVTPEDAG 4228
3981 IVKQDEFTREYCKIGGSPKIVLMYKDETEIQESSKFRMSFVDS-VAVLEMNLVSDESG 4039
4229 --TVSPH--LGNHASSAQLTVRAPEVITILEPLQDVQVLSGQDASQCRLSRAGSQBARWA 4284
4040 DYTCEAHNAGASASSTSLKVEPPIFRKPP-HPIETLKGADVHLECELOQTFFPHVSNY 4098
4285 LGGVPLQANEMNDITVEQ--GTLHLTLHKVTL-----EDAGTVSFHVTCSCSEAO 4334
4099 KDKRELRSKKYKIMSENFELTSHILNVAADIGEQCKATNDVGS-----DTCVGSIAL 4153
4335 KVTAKNTVYRGLNVZALGGHEALFECQLSQEPVAAHTWLLD-DRPVPTSENAEVVFEEN 4393
4154 K--APRFRVKKUSDITVTVGKEVQLOTTIEGAEPIVSVWFKOGEIVRESDNIMWISYEN 4211
4394 GLRHLLLLKNLRPDQ---SCRVTFLAGDMVTSFAFLTVRGWRLEILEPL-----KNAAV 4443
4212 --IATLQFSRVEPANAKYTCQIKNDAG--MQSCFAT----LSVLEPATIVEKPESIKV 4262
4444 RAGAQAARFTCTLSEAVPVGSEASVYINGAAVQPDSDWTVTADGSHQALLRSAPHAGE 4503
4263 TTGDTCTLECTVA-GTPELSTKFWKQKEL-TSDNKYKISFFNKVSGSLKIIINVAPSDSGV 4320
4504 VTFACRAV---ASARLTVLGLPDPEDA-----EVVAHSSHTVTLWS 4543
4321 YSEVQNPVUKDCTASLQVSDRTPPSFRKLKETNGLSGSSVMVECKVYGSPPISVSM 4380
4544 AAPMSDGGGLCGYRVEVEKATGQWRLCHELVGPECVVDGLAPGETYRFRVAAPVPGV 4603
4381 ---FHEGN-----EISSGRKYQITLTDNCT--ALATVMLEESDSDGYTCIATNMAG 4426
4604 AGEVPHLPOTVRLAEPPKPYPPQPSAPESQVAAGEDVSLELVVAEAG-EVIMHKGMER 4662
4427 SDE-CSAPLTVR--EPPSFV-----QXPDPMDVLTGINTVFTTSIVUKTTPPFSVSWFKGSSE 4479
4663 IQPGREPEVVSQGRQWMLVKGTAEQDQGYHGLAQGSGICPAAATFOVALSPASVDEAP 4722
4480 LVFGDRCNVSLDSVLAELFDVDTSGSGEYTCIVS-----NEAG 4519
4723 QPSLPPEAAOEGDHLHLEWALAR--KRMRSREPTLDSISELPEDGRSQRLOPQEAEEVAP 4780
4520 KASC-----TTHLYIKAPAKFVRK-----LNDYSIEKCK-----P 4549

4781 DLSEGYSTADELARTGDADLSHTSSDDESAGTFSLVTVLYLKAGRPOTSPLASKVGPAA 4840
4550 LILEGTF-----GTFPISVTKONG----- 4570
4841 PSVKPOQOQBPPLAARPPGLDSTKOLGDPMSDKAAVKIOAAGKGVKVRKEMKQOQPMF 4900
4571 INVTPSORCNITTEKSPILEIPSTVEDAGQNCVIE-NASGKD-SCSAQILILEPPYPF 4628
4901 SHTFGDTAEAOVGDALELCVAVASKADVRARWLKDGVEL---TDGRHHIQLQDGTCSLL 4957
4629 VKOLEPVKVSVGDSASLQCOLAGTPEIGVSWYKGTKLRTTTYKML-----FRNVATLV 4684
4958 IAGLDRADAGCYTCQVSNKFGQVTHSCVYVVSSESEAESESSGGELDDAFRRARLHRL 5017
4685 ENQVDINDSGEYICKAENSVEGVSASTFLTVOEQ-----KLPPSESRQLRDVQE- 4733
5018 PRIKSPAEDVEBELFLSADGEPAPPEADWQYREDEHEFICIRFEALTTEARQAVTFQE 5077
4734 -----TVGLPVVFDCAISG-----SEPISVSWYKOG-----KPLKDSNNVQTSFLD 4774
5078 MFATLGI-GVEIKLVEGPRRVMEMCISKETPAPVPEPEPLPSLLTSDAAPVFTLQNOE 5136
4775 NTATLNFKTDRSLAGO-----YSCTA-----TNPIGS-ASSARLILTEGNPP 4818
5137 VQD-----GYPVSDCVVTGQPMPSVRHFKDGKLLBEDDHYMINEDQOQGHQJLI 5186
4819 PFDIRLAPDAVVGESADFECHVTGTQPIKVSMAKDSREIRSGKYQISYLENSAH-LTV 4877
5187 TAVVPADMGVYVCLAEHNSMGVSTKAEPLRVLDLSTDYDTAATDESSSYFSAQYLSRE 5246
4878 LKVDKSGQYTCVAVNEVKDCTAQLNI----- 4909
5247 QEGTESTTSGQLPQVVEBELDLQVAPGTRELAKFQL--KVGYPAPRLYKFGQOQPLTAS 5304
4910 RLIPPSFTK--RLSETVEE-----TEGNSFKLEGRVAGSQPITVAVYKNNIEQPT 4958
5305 AHIENTGKILHTLEIISVTRDSGOYAAVYSINAGAAYSARLLVRGPEPEEKPSADV 5364
4959 SNCEITFKNTVLVQVAKAGNDAGLYTCKVNSDAGSALCTSSIVIK--EPKAPPVFDQ 5015
5365 HEQLVPPRMLERFTPKVKKGSSITFSVKVEGVPVTVHVLREAEARGVLWIGP-DTPGY 5423
5016 H-----LTPVTVSEGEYVQLSCHVQSGEPIRIQMLKAGRE-----IKPSDRCSF 5059
5424 TVASAAQSHSLVLLDVGROHOGTYTCTASNAAGQALCASLHVSGLPKVBEQKVKKEALI 5483
5060 SFASGTA--VLELRDVAKADSGDYVCKASNVAGSDTTKSKVTIKDKPAPAPATKKAADV 5117
5484 STFLQGTTOATSAQGLETSF-ADLGGORKEEPLA-AKEALGHLSLAEVGTETEBFLQKLS 5541
5118 RLFFVSPQSRVVEKTTATFIKAVGG---DPIPNVYKWKGRQLNQGRVPIHQ--- 5170
5542 QITEMVSAKITQAKLVQVGGSDSDSKTPS-----ASPRHGRSRPSSSIQ---ESSSES 5592
5171 -----KGDEAKLEI-----RDTTKTDSGLYRCVAFNBHEGETESNVNQLQVDERKKQEK 5217
5593 EDGDARGEIPDIYVVTADYPLGAEQDAITLREGQVVEVLDAAPHLRWLVTATKTPKS 5652
5218 IEGDLRA-----MLUKKTPILKKGAGE-----EEDINELLKNVDP----- 5253
5653 RQGWVSPAYLDRRLKLSPEWGAEEAPFPGAEVSEDEYKARLSVIVIELSSQAP--VE 5710
5254 -----KEYEKYARMYG-ITDFRGLLQAFELLK 5279
5711 ELQFLOSHLOHLERCHPVPIAVAGQKAVIFRNVFDIGRHSFLQELQOCDTDDVAVC 5770
5280 QSQEEETHRLS-----IEEIERSEDO----- 5300
5771 PIKQAAPEQYLEFLVGRVQVQAEVWVSTAIQFYKYKAAEALLAGDSQPPPPLOHYLE 5830
5301 -----EKEFEELVSTIQEL-----SQTEP----- 5320

QY 5831 QPVERVQYQALLKELIRNARNRQNCALLBOAYAVVVSALPORAENKLVHLSMENYPGTL 5890
 Db 5321 -----VTLIKD-IENQTVLKONDAVFE-----IDIKINYP----- 5349
 QY 5891 EALGEPHQHFTWEGAPCARMPKGNRHVFLFRNHLVICRPRDRSRTDTVSVFRNM 5950
 Db 5350 -----E 5350
 QY 5951 MKLSSIDLNDQVEGDDRAPEVQNERDSVRKYLQARTAIKSW-VKEICGQIQRALP 6009
 Db 5351 IKLSWYKTELEPSDK-PEISIDGD-----RHLRVQNCQLKXQGNVLCVCGPHIASAKL 5405
 QY 6010 VWRPDPFEBELADCTAELGETVKLACRVT--GTPKPVISWYKCKAVQVDPHLLIEDP 6066
 Db 5406 TWIEPAWHERHLODVTLEKQGT-----CTMTVQSVNVKSEVFRNGRIILKQGRH-KTEVE 5460
 QY 6067 DGSCALILDSLTGVDGQYMCFAAGNCSTLGLKLVQVPP-REFNVKVRASPFVEGEDA 6125
 Db 5461 HKVHKLTITADVRAEDQGYTC-----KYEDLETSABLRTEAEPIQFTKRIQNLVSEHQA 5516
 QY 6126 QFTCTIEGAPYQIERYKDGALLTTGNKFOTLSEPRSGLLVIVIRAAASKEDLGLYECELV 6185
 Db 5517 TECEV-SDDAIVTWYKGTBELTESQYFNFRNDGRCH--YMTIHNVTPDDEGVY--SVI 5571
 QY 6186 NRL--GSARASAEIRISPMLOAQEQCHREQLVAIVEDTTLERADQVTSVLKELLGPK 6242
 Db 5572 ARLEPRGEARSTAEVYL-----VTKAVKDAKV----- 5599
 QY 6243 ACPSTGDLTGCP-CRGPALQETGQPPVVTGTSSEAPVPPRVPQPLLHGGPQEPPEA 6301
 Db 5600 ---PDIDSRVPIPTMPTRAVPEBI---PPVV-----APPVLLFTP-----BEKKPP- 5643
 QY 6302 IARAQEWTPVIRMEGAAMPGAGTGELLMDVHSHVVRETTORTYTYQAITDHTAAPPMSQV 6361
 Db 5644 ---PKIE-----VTKAVKDAKV-----VAKPEM-- 5668
 QY 6362 TIEDVOAQGTGAQFAIEIGDPQBSVTWYKDSVOLVDSTLSQOQETTVSLVLR--HV 6419
 Db 5669 ---TPREEIVKPPPTTILIPAKAPEIID---VSSKAESVKIMTITRKEV 5713
 QY 6420 ASKDAQVYTCLAQNTGGVCLCAELLVLGGDNEP--DSEKSHRKLHS-FYEVEKEIGR 6476
 Db 5714 QKEAVY-----EKQAVHKEKVFIESPEEYDELEVYTPPEFPQYEPDDEYE 5767
 QY 6477 GVGFVKRVQHGKNIKILCAAKFIPILRSRTAQAYRERDILAALSHPLVTGLDQFETRT 6536
 Db 5768 EIKVEAKKEVHEWE-----EDPEEGQYVERE-----EGYD--- 5799
 QY 6537 LILILELCSBELDLRYKGVVTAEVKVIQQLVEGLHLSH-----GVHLHLIKP 6590
 Db 5800 -----EGEEWEBAQCEVIOQKEVYEE-----SHERKVPKVPKPKAPP 5841
 QY 6591 SNILMVHPAREDKICDFGAQNIIPAELOFQSOYGSPEFVSPEITQQNPVSEASDIWAG 6650
 Db 5842 PPKVIKPVKIE---KTSRMEEEKVQTKV--PE-VSKIVPQKP----- 5883
 QY 6651 VIYLSLTCSSFPAGESDRATLLNVLEGRVSWSPMAHLSEDAKDFIKATLQAPQARP 6710
 Db 5884 -----SRTVQEE-----VIEVKVPAVHTKGMVIEEWMFFASHTEEVSVTVP 5927
 QY 6711 SAAQCLSHPWFLKSPABEAHFINTKQLKLLARSWQSLMSYKSLVMSRISPELLRGP 6770
 Db 5928 EV-----QKEIVTEKHV--AVSKRVEPP 5950
 QY 6771 PDPSLGVARHLCRTDGTSSSSSSSSSSSDELAPFAKSL-PPSPVT----- 6815
 Db 5951 PKVPEL-----PEKPAPEVAPVPIPKVVEPAPKVPKVPKVPKVPKVPKVPKVP 5995
 QY 6816 -----HSPLLHPRGLRPSASLPEAEASERSTAPAPASPEGAGPPAAQGCVPVRSVI 6870
 Db 5996 VPVPKKEAPAPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6050
 QY 6871 RSLFVHQAGESPEHGA-----LAPGSRHRPARRHLLKGYIAGALPGLUREPLMEHRV- 6923

Db 6051 --TIVTQRESPPPAVPEIPKKXVPBEKFPVPRKEEVEVPPPPKVPALP--KKEVPPEKVA 6106
 QY 6924 -----LEEEAAREEQATLLAKAPSFETALRLPASGTHLAPGH--SHSLH 6966
 Db 6107 VPVPVAKKAPPPRAEYVKTKVVEEKFAVEEKLSPVQVEVTRIEVEAEBSWSEEE 6166
 QY 6967 DSRP-----TPRPSSACGE-----AQRUPSPSGGAPTRDMG 6999
 Db 6167 EGVSIYREEEEREAEAEVTEYVVEEPEYVVEEKLHIISKRYEAPPA-----EVT 6220
 QY 7000 HPGSK-----QLPSTGGHPTAQ--PERSPDPSMGQ--PAPFCHPKQGSAPQEGCSHPA 7052
 Db 6221 ERQEKIVLPKIPAKIEPPPPAKVPEAKVIVPEKKVPAPV--PKKVPKVPKVPKVP 6278
 QY 7053 VAPCPGSPFPKCK--EAPL-----VPSSPFLGQOAPAPAPAKASPLDSKMG- 7099
 Db 6279 -KVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6337
 QY 7100 PGDISLPGRPKPCSPGCSASOASSQYSSLVGSSQVGTBPGPSLDAEGHTQE--AED 7157
 Db 6338 PEEKVLKPKRBE--BPPKAVTEFRKRVVKEEKVSTEAPKBPQPIKEVTIMEEKERAYT 6396
 QY 7158 LSDSTTLQRPQOATWRKFSLGGRGYAGVAGYGTFAFGDAGGMLGQGMWARIAMAV 7217
 Db 6397 LEEBAVSQREEEVEEYDYKEFEYEPTEDQY----- 6433
 QY 7218 SQSEEBQSEARABSOEQOARABSPLPQVSARVPVEGRAP-----TRSPSE--PT 7269
 Db 6434 --EYEREYEREEHEEYITEPKIPVKVPPEEVPPTKPAKPAKPAKPAKPAKPAKPA 6491
 QY 7270 PWEDIGQVSLVQIRLDSGDAEAADTISLDSVDPAYLNLSDLYDIKYLPPPEMFIRKVP 7329
 Db 6492 P----- 6494
 QY 7330 KSAQPPSPMAEELAEPEPTWMPGELGPHAGL-----EITESEEDV-----DALLA 7379
 Db 6495 KKLKPPPK-VPEPKVPFEKI-----HISITKREKEQVTEPAKVPKPKRVA 6544
 QY 7380 EAAVGRKRWSPSRSLFHPGR--HLPLD-BPAELGLRERVKASVEHLSRL----- 7429
 Db 6545 EEKVPVRKEVAP-----PVRVPEVPEKEPEVEAEVEV---VTHVEEVLVEEVEEY 6594
 QY 7430 -----KGRPEGEKEGPPPKKP 7446
 Db 6595 IHEEEEFITEEEVVPVIVKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6644

RESULT 12

Q8NHNO PRELIMINARY; PRT; 1020 AA.
 ID Q8NHNO AC Q8NHNO;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Obscurin (Fragment).
 GN OBSCN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Young P., Ehler E., Gautel M.;
 RT "Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 DR EMBL; AJ314908; CAC85755.1;
 DR GO; GO:0005524; F-ATP binding; IEA.
 DR GO; GO:0004812; F-ATPase activity; IEA.
 DR GO; GO:0006418; P-ATPase activity; IEA.
 DR InterPro; IPR003599; Ig.

Qy 6287 -POPLHGEF-----QEPAIARAQEWTPVIRMEGAAM--PGAGTG 6325
 Db 720 GPQPKWVEETIEVRVKMGPOGVSPTE--VPRSSGHLFTLPATPG 767

RESULT 13

Q7Z120 PRELIMINARY; PRT; 8081 AA.

AC Q7Z120;
 DT 01-OCT-2003 (T-EMBLrel. 25, Created)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Uncoordinated protein 89, isoform b.
 GN UNC-89.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=90069613; PubMed=9851916;
 RA Wilson R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z., Le T.T., Wilson R.;
 RT "The sequence of C. elegans cosmid C09D1.";
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Wilson R.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF003131; AAP68958.1;
 SQ SEQUENCE 8081 AA; 894245 MW; 67C804953CF62228_CRC64;

Query Match 7.7%; Score 3182.5; DB 5; Length 8081;
 Best Local Similarity 21.1%; Pred. No. 1.1e-160;
 Matches 1891; Conservative 1190; Mismatches 3350; Indels 2549; Gaps 375;

Qy 27 ATLSQIVGNFTPOVSWKQDPVTAGARFLAQDGD-LYRLTILDALGDSQYVCRAR 85
 Db 565 AKMCLKVTGYPLDITWYKDDVQLHEDERTFYSDDEGDFAMTIDPVQVTDTCRYTCMAT 624

Qy 86 NAIGEAFAAV---GLQYDAEACAEQAPHLRLPTSIRVREGSEATFRCRVGSGSPRAVS 142
 Db 625 NEYGOASTSAFFRVLKVEKEA---PPAFVTKLRDECKEGDVDFCEVEGEWPEPELV 680

Qy 143 WSKDGRILGEPDGRVPRVEELGEASALRIRAAAPRDGGTYEVRANPLGASAAAAALVVD 202
 Db 681 WLVDQDPLRPSHDFRLQVD--GQPAKLEIRDAQDDTGVTYVKIQNEFGSIESKAELFVQ 738

Qy 203 SDAADTASRPTSTAAALLAHORREARAEAGAPSPSTGCTCTCTVTEGCHARLSCVIT 262
 Db 739 ADPPKNHVP-----EFQATIEYVECD-----EGEEVRFKSVIT 772

Qy 263 GERKPEWKKDGLQVTEGRR-HVVYEDAQENFVKILFCQSDR---GLYTCTASNLVG 318
 Db 773 GDPNPEIIFWINGKPLSEKVKFISDGD-----ICLTIKDVTFRHFGDMVTCQGSNRLG 827

Qy 319 QYSSVLVVRPAVP--FKRLQDLQLEVRKESATFLCEV---POPSTEAAWFKBEETLW 373
 Db 828 SASCDGRKVRVPPAPPTFNKPLEDKTVQEKSTVVFYDVSGWPEPTLTFTLCKELK-- 885

DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003598; IG_c2.
 DR InterPro; IPR001849; PH-
 DR InterPro; IPR000219; RhGEF.
 DR InterPro; IPR001452; SH3.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00047; IG; 2.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00621; RhGEF; 1.
 DR SMART; SM00409; IG; 2.
 DR SMART; SM00408; IGC2; 2.
 DR SMART; SM00235; PH; 1.
 DR SMART; SM00325; RhGEF; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 DR PROSITE; PS0010; DH_2; 1.
 DR PROSITE; PS00835; IG LIKE; 2.
 DR PROSITE; PS00003; PH DOMAIN; 1.
 KW Immunoglobulin domain.
 FT NON_TER 1
 SQ SEQUENCE 1020 AA; 112239 MW; ALFABD68B3F1948B_CRC64;

Query Match 7.9%; Score 3270; DB 4; Length 1020;
 Best Local Similarity 84.9%; Pred. No. 6.4e-167;
 Matches 654; Conservative 14; Mismatches 54; Indels 48; Gaps 9;

Qy 5601 IFDIYVVTADYLPGLAEQDAITLREGQYVEVLDAAHPLRLVTRTKTSSPSRQGWSPA 5660
 Db 1 IFDIYVVTADYLPGLAEQDAITLREGQYVEVLDAAHPLRLVTRTKTSSPSRQGWSPA 60

Qy 5661 YLDRRLKLPENGAAPFPPEGAVSEDEYKARLSVIOELSSQCAFVEELQFLQSHHL 5720
 Db 61 YLDRRLKLPENGAAPFPPEGAVSEDEYKARLSVIOELSSQCAFVEELQFLQSHHL 120

Qy 5721 QHLRCPHYPIAVAGKAVIFNRVDRIGRHFSSFTQELQCCDTDDVAMCFIKNOAFAEQ 5780
 Db 121 QHLRCPHYPIAVAGKAVIFNRVDRIGRHFSSFTQELQCCDTDDVAMCFIKNOAFAEQ 180

Qy 5781 YLEFLVGRVQAEVSTVTAIOEFYKYABEALLAGDPSPPPPLQHYLEQVVERVQYQ 5840
 Db 181 YLEFLVGRVQAEVSTVTAIOEFYKYABEALLAGDPSPPPPLQHYLEQVVERVQYQ 240

Qy 5841 ALLKELIRKARNRQNCALLEQAYAVSALPQRAENKLVSLMENYPGTLEALGPPIRQG 5900
 Db 241 ALLKELIRKARNRQNCALLEQAYAVSALPQRAENKLVSLMENYPGTLEALGPPIRQG 300

Qy 5901 HFTVWEGAPGAPWGHNRHVPFLFNHLVICKPRDSRTDTVSVYFRNMKLSLIDLND 5960
 Db 301 HFTVWEGAPGAPWGHNRHVPFLFNHLVICKPRDSRTDTVSVYFRNMKLSLIDLND 360

Qy 5961 QVEGDDRAPEVWQEREDSVRKYLLOARTAIKSSWYKEICGQORLALPVRPPDFEEL 6020
 Db 361 QVEGDDRAPEVWQEREDSVRKYLLOARTAIKSSWYKEICGQORLALPVRPPDFEEL 420

Qy 6021 ADCTAELGETVKLACRVGTGPKPVIWYKDGKAVQVDPHILIEDPDGSCALILSLTGV 6080
 Db 421 ADCTAELGETVKLACRVGTGPKPVIWYKDGKAVQVDPHILIEDPDGSCALILSLTGV 480

Qy 6081 DSGQYMCFAASAGNSTGLKILVQVPPFVNVKVRASPFVEGDAQFTCTIEGAPYQIR 6140
 Db 481 DSGQYMCFAASAGNSTGLKILVQVPPFVNVKVRASPFVEGDAQFTCTIEGAPYQIR 540

Qy 6141 WYKDGALLTGNKFTLSEPRSGLLVIVRAASKEDGLGYECVLNRLGSASAEFLRQ 6200
 Db 541 WYKDGALLTGNKFTLSEPRSGLLVIVRAASKEDGLGYECVLNRLGSASAEFLRQ 600

Qy 6201 SPMLQAEQCHRLQVAAVEDTLER-----ADQEVTSVLKLLGPKACGPSGDLT 6252
 Db 601 SPMLQAEQCHRLQVAAVEDTLER-----ADQEVTSVLKLLGPKACGPSGDLT 659

Qy 6253 GPGPCPRG-----APAL-----QETGSP--PVTGTSEAPVPPRV--- 6286
 Db 660 SPSPCSASPLRPGLLAPDLLLPAGACGPRRPEAPGKPVWVPLTYTEAEASPALPGLS 719

| | | | |
|----|------|--|------|
| Db | 2591 | TFPVKMYICTQLVCATEDTEISNANGVY-TMKIHGVTADMTGKIKCVAYNKAGEVSTBG- | 2648 |
| Qy | 2372 | FFVESQSIIVRGLOQVTVMEPAPAMFE---CETSI-----PSVRPPK | 2411 |
| Db | 2649 | -----PLKVAPIPVEFETSLCDATCREGDTLKLRAVLLGPEPVW---S | 2690 |
| Qy | 2412 | WLLGKTVLQAGNVGLEBGTVHRLMLRRTCTSTMTPGVHFTV---GKSRSASRLVWSDI | 2467 |
| Db | 2691 | WTVNGKKLEESQNIKIHSEKGYTYTWIKDICTDSGVQVCEALNEYGKATSEATL---- | 2745 |
| Qy | 2469 | PVVLTRPLEP-----KTG-RELQSVVLSCDFRPAKPAVOWYKDDTFLSPSEKFK | 2515 |
| Db | 2746 | -LVLPRGPPDFLEWLSNVRAARTGTWVHVFTGDPKP---SLTWYNNKEILNSDLYT | 2801 |
| Qy | 2516 | MSLEQMAELRLRLMP-----ADAGVVRCOA-----GSAHSSTEVTEARE | 2557 |
| Db | 2802 | IVTDDKTSLTTINSENPVHVGEIICKAENDAGEVSCANMITYTSDMFSSESBAQAE | 2861 |
| Qy | 2558 | -----VTVTGP-----LQDAEATEEGWASFSCELSHEDES---VE | 2589 |
| Db | 2862 | FVGDDLTDESUREEMHRTPTVPMAKPTIKDKTKAKKGHSAVFCVV---PDTKGVCK | 2919 |
| Qy | 2590 | WSLNGMPLYNDSFHISHKGRHT-----LVLSIORADAG-----IVRASSLVK | 2634 |
| Db | 2920 | WLKDGKEI-----ELIARIVQTRTGPBGHITQBELVLDNVTPEDAGKYTCIVENTAGD | 2973 |
| Qy | 2635 | STASARLEV-----RVKPVVFLKALDLSAEBRGTTALQCEV-SDPEAHVVRKDGQV | 2685 |
| Db | 2974 | TCEATLTVIESLEKSEKAPFIVALQDKTKTSEKVVLECKVIGEPKPKVSMLDNKT | 3033 |
| Qy | 2686 | LGPSDKYDLHAGTRGLVHVDVSPEDAGLYTCHV-----GSEETRARVRHDLHVGITRK | 2741 |
| Db | 3034 | I-TQESITVESVEGVERVITTSLSHQGKYTCIAENTEGTSKTEAFLTVQGEAPVFYFKE | 3092 |
| Qy | 2742 | LXTMEVLEGESCFEVLVSHESASDPAMVTGCKT-----VGSSSRF-----QATROGRK | 2791 |
| Db | 3093 | LQNKELSIGEKLVLSCVSGSQPHVDFYSFSETTKVETKINTSSRIALIEHDQTHWM | 3152 |
| Qy | 2792 | YI-LVVREAPSDAGEWFSVRGLTSKASLIYRERPAALIKPLEDQWVAPGEDVELREL | 2850 |
| Db | 3153 | VISQITKEDIVSKALATNSIGTATSTSKITTKVEAPVEQGLKKTSTVKEKEIKWEVKV | 3212 |
| Qy | 2851 | SRAGTPVHVKOKAIRKOKYDVV---CEGTAMVIRGASLKADAGEYTCVEASEKSTAS | 2908 |
| Db | 3213 | GGAPVDFWKDDPKPSEGNHEMKNPETGFTLVVQAAITDAGKYTAKASNPAGIAE | 3272 |
| Qy | 2909 | LHVEEKAN-----CPTTEL---TNLQVEEKGAVETCKTEHPAATVWRKGLLELRASGK | 2960 |
| Db | 3273 | SSAAEAVTQSLKPTVRELVTTEVKINETATLSVTVKGV-FDPSVEMLKQDQPVQTDSS | 3331 |
| Qy | 2961 | HQPSQ-EGT-TLRLTISALEKADSDTYTC-----DIGQASRAQLLVQGRV--HIIEBLE | 3012 |
| Db | 3332 | HVIKVEGSGYSITIKDARLEDSGKYACRATNPAGEAKTEANFAVVKNLVPPFVEKLS | 3391 |
| Qy | 3013 | DVDVQEGSGATFCRISIPANYEPVHFLDKTPLHANELNEIDAQPG-GYHVLTLRLQALK | 3071 |
| Db | 3392 | PLSVKEKESTLSVKVGVGPEPSEVNEFKDDTFISIDNVHVIQKQTAGVSFSLTINDARQ | 3451 |
| Qy | 3072 | DSGTIYF-----BAGQORASALURVTEK--PSVFSRELDTATITEGEDTLVLCSTCDI | 3124 |
| Db | 3452 | DVG-IYSCRARNEAGEALATNPAGIIRDSIPPEFTOKLRPLEVREQETLTLKVTVIGTPV | 3510 |
| Qy | 3125 | P-MCWNKDKGTLRGSARCQLSH-----EGHBAQLITGATLQDSGRYKC---EAGGA | 3172 |
| Db | 3511 | PNVWFKDDKPI-----NIDNSHIFAKDEBGGHIT-ITIQARGEDGVVTCATNEAGEA | 3565 |
| Qy | 3173 | CSSSIVRVHAR---PVRFOBALKDLEVLGGAAATLRCVLSVAAP-VKWCYGNVNLRLPGD | 3228 |
| Db | 3566 | KTTANNAVQBEIEAFL-FVQGLKPYEVEQGPALVVRVEGKPEPEKPKWFKDGVPIADN | 3624 |
| Qy | 3229 | KYSLRQBG---AMELVVRNLRLPODSGRYSC-----SFGDQTTSATL-----TVTAL | 3272 |
| Db | 3625 | QHVIERKKGSGSHTLVIKDTNNADFGKYTCQATNKAGKOBTVGELKIPKYSFEKQABEV | 3684 |
| Qy | 3273 | PAQFIGKLRNKEATEGATATLRCELSKTA--PVWEKSGSETLRDGDYRCLR--QDGAMCE | 3328 |
| Db | 3685 | KPIFIEPLKETFAVEGDTVVLECKVNKESHPPQIKFFKNDQPVIEIGHQMLVEDGNI-K | 3743 |
| Qy | 3329 | LQIRGLAMWDAABYSC-----VCGEERTSASLTI----- | 3357 |
| Db | 3744 | LTIQNAKKEDVGAYRCEAVNVAGKANTNADLKIQFAAKVEEHVTDSSGQLEEQGFETVG | 3803 |
| Qy | 3358 | -----RPPAFHFIGRLRHOESIEGATATLRCELSKAAAP---VWEK-KRESLRDG | 3403 |
| Db | 3804 | DTASSKTDGORGAP-BFVELLRSCVTTEKQOAILKCKV-KGEPRPKIKWTKKEKEVMSA | 3861 |
| Qy | 3404 | DRHSRLRODGAVCBELQICGLAVADAGYSCVCGBERTSA-----TLTVKALP-----AK | 3451 |
| Db | 3862 | RVRAEHKDDCTLTLTDFNVTOADAGYRCEAENEYGSAMTEGPITVLEGAPKIDGEAPD | 3921 |
| Qy | 3452 | FTEGLRNEEAVEGATAMWLCELS-KVAP-VWEKSGPENLRDGDYRILR--QESTRCELQI | 3507 |
| Db | 3922 | FLOPVKPAVVTVGETAVLEGIKSGKPKSVKYNKEELKPSDRVKIENLDDGTQ-RLTV | 3980 |
| Qy | 3508 | CGLAMADAGEYLCVCGQE-----RTSATLITRALPAR-----FIEDVKNEAREGATAVLQ | 3559 |
| Db | 3981 | TNAKLDMDDEYRCEASNEFGVMSDVTLTVKE-PAQVAPGFFKELSAIQVKETETAKFEC | 4039 |
| Qy | 3560 | ELNSAAP-VWEKSGSETLRDGD--YSLRODGTKECELOIRGLAMADTGEY---SCVCGQ | 3612 |
| Db | 4040 | KVSGTKPDVKWFKDGTPLKEDKRVHPFESTDDGTQ-RLVIEDSKTDDQGNTRYIEVSDNAGV | 4098 |
| Qy | 3613 | ERTSAMLTVRALP---IKFTPEGLRNEAREGATAVLRCEL-SKMAPVEMWKGHETLRDGD | 3668 |
| Db | 4099 | ANSKVPITV--VPSSETLKIKKGLTDVNVTOGTKILLSVEVEGPKTKVKYKGTETVTSQ | 4156 |
| Qy | 3669 | RHSLRQ-DGARCELOIRGLVAEDAGEYLCMCXKE---RTSAMLTVR-----AMPKFI | 3717 |
| Db | 4157 | TTKIVQVTESEYKLEIESAEMSDTGAYRVVLSTDSFVSSESAITVTVKAAEKISLPS-FK | 4215 |
| Qy | 3718 | EGLRNEAREGDTATWLCEL-SKAAAPVWEKKGHETLRDGDHSLRQDGSCELOIRGLAV | 3776 |
| Db | 4216 | KGLADQVPKGPPLVLEVEIEGPKDKVKYKNGDEIKDKGVEDL--GNGKYLTIPIPFQE | 4273 |
| Qy | 3777 | VDAGEYSCV---CGQERTSATLTVRALPARFIEDVKNQAEAREGATAVLQCEL-SKAAPV | 3831 |
| Db | 4274 | KDVEYSVTAANEAGEIESAKVNVSAKP-EIVSGLVPTTVKQGETATFNVKVKGPVKGV | 4332 |
| Qy | 3832 | EWKSGSETL-----RGGDRYSRLRODGTCELOIHGLSVADTGEYSCV---CGQERTS | 3880 |
| Db | 4333 | KWYKNGKEIPDAKTKDNGDGSYS-----LEIPNAQVEDAADYKVVVNSNDAGDADSS | 4393 |
| Qy | 3881 | ATLTVRAP-----QPVFREPLQSLQAEBSGSTATLOCELSSEPTATVWVSKGGLQLOANG | 3933 |
| Db | 4384 | AALTVKLADGDKVKKEIIVSGLIPTTVKQGETATFNVKVKGPVKQVKYKNGKEI-PNA | 4442 |
| Qy | 3934 | RREPRLOGCTAELVLQLOREDTGEYTCYTCGSA-----TSATLTVTAAPVRFLELOHCE | 3989 |
| Db | 4443 | KAKDNGDG-SYSLEIFPNAQLDDTADYKVVVNSNDAGDADSSAALTVPKLPGLIATVKGLEDAE | 4501 |
| Qy | 3990 | VDEGTAHLCCCELGRAGASVWEKSGSLQLPFCAYQMVQDG-AAAEALLVRGVEQEDAGDY | 4048 |
| Db | 4502 | VPKKKAVLVQETWKKPKKEIKWYKNGKEIIPSDKAQFGSDGDNKFPOLVIPDAGDDDAEY | 4561 |
| Qy | 4049 | -----TCDTGHT-QSMASLSVRVP--RPKFKTRLOSLEQETGDI-ARLCCQLSDAESAVV | 4100 |
| Db | 4562 | KWLTDDEGNTADSSCALTVKLPAKEPKI---IKGLEDDQVVSIGSPKLEIETSGSPKT | 4618 |
| Qy | 4101 | QWLKEGVELHAGPKVVRSGQATRELLIHQLEA---KDTGEYACVTGQKTA--SLRVT | 4155 |
| Db | 4619 | KWYKNGELPGLAAKTIKQIKIDNKKVLEIPSSVEEDGTGYKVEVANEAGSANSKKIT | 4678 |
| Qy | 4156 | -EPEVTIVRGIVDAEVTADEVFEFSCVSRAGATGVQWCLQGLFQSNVEVTEAVRDGRI | 4214 |
| Db | 4679 | VEPKITFLKPLKQDSIIEGENAEAFSEVINTKPRI-VKWKYKNGQBIKPNRSFIIIEQKDTDK | 4737 |

Db 6261 TBEYGFKNLNTASLPTTDRGPFKEVTHGYLTLWIFTKAPRYPQVTVIIEIRLPE 6320
QY 6263 ---ALQETGSPPTV---GTSE-APAVPP---RVPOPLLHE 6293
Db 6321 KQMSLLEYNIPVCKVRNLELGSQYQVRVRAENIYIGISDPSASPRLMAPQPVDR 6380
QY 6294 ---GPOPEPEALARAQE---W---TVPIRMGE- 6316
Db 6381 RINKVILPLDPAEAKALDMRYSEQYACAPWFSGVYKRYCAENDTLTILVNVSGFPDPD 6440
QY 6317 AAWPGAGTCELLWDVHSHVRETTORT---YTQOAIIDHTA--- 6354
Db 6441 IKWKFRG---WDI---DTSSTSKCYTYTGGSETTLAITGFSKENVQYQCPAK 6499
QY 6355 ---RPSMQVTIEDVQAGTGAQ---FEALIEGDPQPSVTWYKDSV 6395
Db 6490 NDYGAQOQIMVDLATRNFICPLVN---KTFSSAQPMRMDVRVDGEFFPELKKWKEWR 6545
QY 6396 QLVDSRLSQOQEGT-TYSLVLRHVASKADAGVYTCLAQNTGCVLCKAELLVGLGDNEPD 6454
Db 6546 PIVESSRHKFVQDGPVLCSLIINDPMWRDSGLYSCVAVNDAGQATSTCTVTVEABGDYND 6605
QY 6455 SEKQSHR---RKLHSFYVEKEEIGRGVGFVRVQHKGNKILCAAKFIPILRSRTRAQA 6509
Db 6606 VELPRRRVTIESRRVRELYEISE---KDEKL--AABGAPFRVKEKATG 6648
QY 6510 ---YREDILALSHPLVTGLLDQOFETRTKLILILELCSSEELLDRLYR 6555
Db 6649 REFLAQLRPIDALMRHVHNSLDHPGIVQHRVLRDEKALVVD--NANSTIDGLSS 6706
QY 6556 ---KGVTEAEVKYVIOQLVEGLHLSHGVLHLDIKPNSILMVHPAREDIK 6604
Db 6707 LAHPGVIEAEPKGVNRETCTRVFVQQLLALXEMHDLRAHLDRPEYILL---QDDKLK 6763
QY 6605 ICDFGAQNTPAELQFOSGPFVSPBEIIQONPVSEASDTWANGVYSYLSLTCSGSPA 6664
Db 6764 LADFGQARLLRLGITGEIKGSPFVSPBEIVRSYPLTLATDMWSTGLVTLVLLTGLSPFH 6823
QY 6665 GSDRATLNVLEGRWSSWAHLSEDAKDFIKATIQAPAPRPSAQCILSHWPF--- 6721
Db 6824 GDNDETLANVSCQPD-SSPL-GNFSYDAGDFVXKLITEIPVSLTVDREALDHPWIND 6881
QY 6722 -LKSMFAEBAHFTINTKQLFLLARSWQSLMSYKSLVMRS---IPELLRGP- 6770
Db 6882 KLKTEP-LSADTLREFYQHKWLE---RRVFQQTSPSEQLLEAILGPATAQAQ 6930
QY 6771 PDSPLGLVARHLR-----DTGSSSSSSSDNE- 6799
Db 6931 QNAPVAPEGRRPABIYDYLRIQPKXPPPTVYVQPRKEHPFFIDFEGOLIDGADFADPE 6990
QY 6800 ---LAPFARAKSLPPSP---VTHS-----PLLHPRGFL 6826
Db 6991 GTCFEGPHRQPPIPPQOPQPNQAHDSRRHQOQPOHQOQORIPVDQYGRPLVDPR-YL 7049
QY 6827 RPSASLPESAERATEAPAPSPAGAGPPAAQCQVPRHVSIRLSLFYHQAGESPEHGA 6886
Db 7050 NDSHRPSSLDADFFYVDKYNPVHFDKYGRPMA---PQNLKXKLIPQDKGETFSSH- 7104
QY 6887 LAFGSRHRPARRHLLKGGVIAGALPG-----LRPLMEHRVLEBEAAR-----E 6931
Db 7105 -KKEKCHPVATPIL-----ASFGQOQOQKIPMRMIRGERIEIEIANRILSDIS 7155
QY 6932 EQATLLAKAPSPETALRLPASGTHLAPGSHSHLEHDSPTTPSPSACGEAQRPLSPSG 6991
Db 7156 EEGSIAGSLASLED-FEIP-----KDFQVEASEPSTPLTPE----- 7191
QY 6992 GABIRDMGHPOGSKQLPSTGGHGTAPQPERPSPDPSWGPAPFCHPKQGSAPQEGCSHPH 7051
Db 7192 -VTIRE-----TIPKPTPTSPOKSPVP----- 7214
QY 7052 AVAPCPFGSPGCKEAPLVPSSPFLGQOPAPPAPAKAS-----PPLDSKM---G 7099
Db 7215 -----OPQGLIIPAKVYSDSILLAG:PAADKKVLEDA 7246

QY 7100 PGDISLP-GRP--KPGPCSPGASQASQSSVSLRAGSQVTEPQPSLDAEGWTQAE 7156
Db 7247 ENDPSPVPGAPLFLGLEHSGDLTIDTTSASGL--IKVTPAINLSNPK-----SPRRS 7298
QY 7157 DLSDSPTLQPOEQATMRKFSLGGGGYAGVAGYGTFAFGDAGMGLGQPMWARIWA 7216
Db 7299 TPGTKSVVLSPQKQSHMEVLIATKRG-----KPGFLPPGEL----- 7335
QY 7217 VSQSEEEQOEAEAEASQSEQOEAEAPLQVQSPVPVEVGRAPTRSSPEPTWEDIGQ 7276
Db 7336 ---AEDIDDEDAFMDRKKQ----- 7352
QY 7277 VSLVQIRLDSGDAEAD-----TISL-DISEVDP--YLNLSDL-----YDIKY 7317
Db 7353 ---VKPDHGDENDFKDEKERLEKDKNRTVNLDLDKYPSPAFYKDDSDFGHPGDIDA 7409
QY 7318 LPFE---FMI---FRKVPKSAQPEPSPMAEBELABFPEPTWPPWBGELG-PHAGLEITEE 7370
Db 7410 TPWDSHVQIGPDYILWAARGAFAFNSVRVNRVRELFGMGAPTVK-QGFLGVNRDITVRER 7468
QY 7371 SEDVDALLAAAVGRKRWSSPSRSIHPFPGRHLPDEPAELGLRLRVRKASVHISRLK 7430
Db 7469 RRYTD-ILRETTOGLEPKSHEQSTAL-----LQAFSATAIERIKADIKVTCP-- 7516
QY 7431 GRPEGLEKEGPPRKPKGLASFRLSGLKSMWDRAPTEFLRELSDETIVLQSQSVTLACQVSAQP 7490
Db 7517 -----ATKNDGTF-----APIFTARLADVLRKNQPAIFECASVSP 7555
QY 7491 AAQATWSKDGAPLESRRVLSIATLKNFQLLTILVVVAEDLGVYTVCSVNALGTVTGGV 7550
Db 7556 APKVTWDFQKILESNDRVITEQD-NNVARLIINHAAPYDLGBYVCTVTAINEYGTDKSSCR 7614
QY 7551 LRKAERSSSPCDIGEVYAD-GVLLVWKPVESVGP-----VTYIVQCSL-----EGSW 7599
Db 7615 LISGETSPRGRPE-AELSSDTEIFIQWEAPE--GPYLEGITYRLRYRVRVAGNDHGDWP 7671
QY 7600 TTLASDIFDCYLTSLKSRGTYTFRACVSKAGMGPYS-----SPSEQV--- 7644
Db 7672 ITVSEKIDDESIVKHLSPGLIYQFVTAQNGFGLGLPSLSRIVQTHGKGAQKLQIDVL 7731
QY 7645 -----LLGPGSHLASEESQGRSA-----QPLSTKTFAQTQ 7677
Db 7732 KSEIRLVNVMPQKSTNQLGISE-ESEEDSEARTANEDMKSNLQOTDDPTGRFQIGGL 7790
QY 7678 IQGRFVSVQOCWEKAS-GR-A-LAAKIIPYHPKDKTAVLREYEAALKGLRHPHQAHAAY 7735
Db 7791 KFKGRFSVIDAVDSTTEGHACAVKI--RHPSE--AISEYSLRDQGHENVQRLIAAF 7846
QY 7736 LSPRHVLIILELCSGPPELLCLAEASYSSESEVKYLMQMLSATQVIHNOHILHLDRSE 7795
Db 7847 NNSNLYLLSERLY-EDVFSRFVNDYTEEQVALTMRQVTSALHFLHFKGIAHLDVNPH 7905
QY 7796 NMIIITEYN--LLKVVLDLGNASLSQEKVLPDCKFQYLET--NAPEL-LEGQAVPOTDI 7850
Db 7906 NIMFQSKRSWVKLVDFGKQKVS-GAVKPD-----FDYKVASPEPHIETPTVTVQSDM 7959
QY 7851 WAIGVTAFIMLSAEYVSVSESGARDLQRLKGLRVLRSRCYAGL-----SGGAVAFRLSTL 7905
Db 7960 WGMGVVTFCLLAGHPFTSE--YDREBEIKENVIN-KCDPNLIPVNASQECLSFATWAL 8016
QY 7906 CAQWGPACPASSCLOCPLTEEGPACSRPAPVTFPAPLAR 7945
Db 8017 KKSVPVRMTDEALSHKFLSDSPSMVRRRESIKYSASRLR 8056

RESULT 14

Q8NH8 PRELIMINARY; PRT; 646 AA.
ID Q8NH8
AC Q8NH8;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Obscurin (Fragment).
GN OBCSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Young P., Ehler E., Gautel M.;
RT "Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere
RT assembly";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314898; CAC85746.1; -.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR008957; FN.III-like.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 4.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00409; IG; 6.
DR SMART; SM00408; IGc2; 4.
DR PROSITE; PS0835; IG LIKE; 4.
KW Immunoglobulin domain.
FT NON_TER 1
FT NON_TER 646
SQ SEQUENCE 646 AA; 70258 MW; B04F5B00E54323D68 CRC64;
Query Match 7.5%; Score 3113; DB 4; Length 646;
Best Local Similarity 93.5%; Pred. No. 8.3e-159;
Matches 605; Conservative 16; Mismatches 24; Indels 2; Gaps 1;
QY 421 NLLRLPKRTAVRGDTAMFCVELAVPGVPHVLRNQEEVAGGRVAISAEGRHTLTIS 480
DB 1 NLLRLPKRTAVRGDTAMFCVELAVPGVPHVLRNQEEVAGGRVAISAEGRHTLTIS 60
QY 481 QCLEDVGQVAFMAGDCOTSTFCVSAPRKPLPPDPVVKARMESSVILSWSPPPHGE 540
DB 61 QCLEDVGQVAFMAGDCOTSTFCVSAPRKPLPPDPVVKARMESSVILSWSPPPHGE 120
QY 541 RPTVIDGYLVEKKLGTGTWIRCHAEWVATPELTVDVAEENGFQFRVSALNSFGQSPY 600
DB 121 RPTVIDGYLVEKKLGTGTWIRCHAEWVATPELTVDVAEENGFQFRVSALNSFGQSPY 180
QY 601 LEFGTGHVLAFLKAVRTPLKAVQAVEGGEVTSVDLTVASGENFLDGOALKASSVYEIH 660
DB 181 LEFGTGHVLAFLKAVRTPLKAVQAVEGGEVTSVDLTVASGENFLDGOALKASSVYEIH 240
QY 661 CDRTHTLTIREVPASLHGAQLKFVANGIESSIRMEVRAAPGLTANKPPAAAAAEVLARL 720
DB 241 CDRTHTLTIREVPASLHGAQLKFVANGIESSIRMEVRAAPGLTANKPPAAAAAEVLARL 300
QY 721 HEEAQLLAELSDQAAAVTWLKDGRTLSPGPKYEQASAGRVLLVRDVARDDAGLYECVS 780
DB 301 HEEAQLLAELSDQAAAVTWLKDGRTLSPGPKYEQASAGRVLLVRDVARDDAGLYECVS 360
QY 781 RGRRTAYQLSVQGLARFLHKDMAGSCVDVAVAGGPAQFECETSEAHVHVHYKDGMELGHS 840
DB 361 RGRRTAYQLSVQGLARFLHKDMAGSCVDVAVAGGPAQFECETSEAHVHVHYKDGMELGHS 420
QY 841 GERFLQEDVGRHRLVAATVTRQDEGTYSYCVGSDSDVDFRLRVSEPKVVFAXEQALARKL 900
DB 421 GERFLQEDVGRHRLVAATVTRQDEGTYSYCVGSDSDVDFRLRVSEPKVVFAXEQALARKL 480
QY 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVCWEATGCTFRLRVQOAGADAGEY 960
DB 481 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVCWEATGCTFRLRVQOAGADAGEY 540
QY 961 SCEAGGQRLSHLDVKEPKVFAKDQVAHSEVQAEAGANATLSCEVAQAQTEVTWYKDGK 1020
DB 541 SCEAGGQRLSHLDVKEPKVFAKDQVAHSEVQAEAGANATLSCEVAQAQTEVTWYKDGK 598

QY 1021 KLSLSLKVHVEAKGCRRLRVVQOAGKTADGYSCEARGQVSRFLHI 1067
DB 599 KLSLSKVCWEATGCTFRLRVVQOAGADAGEYSCGAGGQRLSHLDV 645
RESULT 15
Q8NH7 PRELIMINARY; PRT; 551 AA.
ID Q8NH7
AC Q8NH7;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Obscurin (Fragment).
GN OBCSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Young P., Ehler E., Gautel M.;
RT "Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere
RT assembly";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314900; CAC85747.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00047; ig; 6.
DR SMART; SM00409; IG; 6.
DR SMART; SM00408; IGc2; 6.
DR PROSITE; PS0835; IG LIKE; 6.
KW Immunoglobulin domain.
FT NON_TER 1
FT NON_TER 551
SQ SEQUENCE 551 AA; 59655 MW; 97D8C06CF57F822F CRC64;
Query Match 6.8%; Score 2793; DB 4; Length 551;
Best Local Similarity 99.5%; Pred. No. 1.1e-141;
Matches 548; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 978 PKVFAKDQVAHSEVQAEAGANATLSCEVAQAQTEVTWYKDGKLSKSLKVHVEAKGCR 1037
DB 1 PKVFAKDQVAHSEVQAEAGANATLSCEVAQAQTEVTWYKDGKLSKSLKVHVEAKGCR 60
QY 1038 RLWVQOAGKTADGYSCEARGQVSRFLHIPEKMFVAKESVHNEVQAEAGASAMLSCE 1097
DB 61 RLWVQOAGKTADGYSCEARGQVSRFLHIPEKMFVAKESVHNEVQAEAGASAMLSCE 120
QY 1098 VAQAQTEVTWYKDGKLSKSSKVCWEATGCTFRLRVQOAGADAGEYSCGAGGQVSRFL 1157
DB 121 VAQAQTEVTWYKDGKLSKSSKVCWEATGCTFRLRVQOAGADAGEYSCGAGGQVSRFL 180
QY 1158 HITEPKGVFAKESVHNEVQAEAGTAMLSCEVAQFQTEVTWYKDGKLSKSSKVMVEK 1217
DB 181 HITEPKGVFAKESVHNEVQAEAGTAMLSCEVAQFQTEVTWYKDGKLSKSSKVMVEK 240
QY 1218 GCTRLRVQOAGKADAGEYSCGAGGQVSRFLHIPEKMFVAKESVHNEVQAEAGASAT 1277
DB 241 GCTRLRVQOAGKADAGEYSCGAGGQVSRFLHIPEKMFVAKESVHNEVQAEAGASAT 300
QY 1278 LSCEVAQAQTEVTWYKDGKLSKSSKVCWEATGCTFRLRVQOAGADAGEYTCAGGQRL 1337
DB 301 LSCEVAQAQTEVTWYKDGKLSKSSKVCWEATGCTFRLRVQOAGADAGEYTCAGGQRL 360
QY 1338 SFHLDVSEPKVFAKESVHNEVQAEAGANATLSCEVAQAQTEVTWYKDGKLSKSSKVR 1397
DB 361 SFHLDVSEPKVFAKESVHNEVQAEAGANATLSCEVAQAQTEVTWYKDGKLSKSSKVR 420
QY 1398 MEAVGCTRLRVQOAGADAGEYSCGAGGQVSRFLHIPEKMFVAKESVHNEVQAEAG 1457
DB 421 MEAVGCTRLRVQOAGADAGEYSCGAGGQVSRFLHIPEKMFVAKESVHNEVQAEAG 480

Wed Sep 22 12:33:16 2004

| | | | |
|----|------|--|------|
| Qy | 1458 | ASTTSCVAAQTEVVMYXDGKLSFSSKVRMEAVGCTRRLLVVQAGQADAGEYSCEAG | 1517 |
| Db | 481 | ASTTSCVAAQTEVVMYXDGKLSFSSKVRMEAVGCTRRLLVVQAGQAVAGEYSCEAG | 540 |
| Qy | 1518 | SORLSFHLHVA | 1528 |
| Db | 541 | SORLSFHLHVA | 551 |

Search completed: September 13, 2004, 11:35:51
Job time : 636 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:07:04 ; Search time 2831.1 Seconds
(without alignments)
10242.013 Million cell updates/sec

Title: US-10-077-130-4_COPY_23150_24120

Perfect score: 971

Sequence: 1 cagcggcgccggcgctggcgg.....ggccgctaaaaagtctaga 971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_esti.*

9: gb_esti.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: gb_estfun.*

15: em_estom.*

16: em_gss_hum.*

17: em_gss_inv.*

18: em_gss_pln.*

19: em_gss_prt.*

20: em_gss_fun.*

21: em_gss_fam.*

22: em_gss_mus.*

23: em_gss_pro.*

24: em_gss_rod.*

25: em_gss_phg.*

26: em_gss_vri.*

27: gb_gss1.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| C 1 | 713.4 | 73.5 | 715 | 12 | BM985339 |
| C 2 | 672.2 | 69.2 | 731 | 12 | EG818748 |
| C 3 | 626.4 | 64.5 | 989 | 13 | BX435097 |
| C 4 | 480 | 49.4 | 480 | 13 | BX280322 |

| | | | | | |
|------|-------|------|------|----|----------|
| C 5 | 399.8 | 41.2 | 640 | 14 | CD774776 |
| C 6 | 377.4 | 38.9 | 622 | 14 | CF176441 |
| C 7 | 339 | 34.9 | 768 | 10 | BF134040 |
| C 8 | 331 | 34.1 | 479 | 10 | BF651426 |
| C 9 | 313.4 | 32.3 | 522 | 13 | EQ554402 |
| C 10 | 311.4 | 32.1 | 493 | 10 | BE110731 |
| C 11 | 247.4 | 28.5 | 494 | 14 | CF177813 |
| C 12 | 243.4 | 25.1 | 428 | 9 | AI503993 |
| C 13 | 238.2 | 24.5 | 408 | 10 | BF414891 |
| C 14 | 235.2 | 24.2 | 418 | 13 | EX636992 |
| C 15 | 231.6 | 23.9 | 445 | 13 | BY588786 |
| C 16 | 227.6 | 23.4 | 411 | 12 | BI186010 |
| C 17 | 214.8 | 22.1 | 494 | 14 | CB713368 |
| C 18 | 209.8 | 21.6 | 417 | 13 | BY437537 |
| C 19 | 209.6 | 21.6 | 379 | 10 | BE754373 |
| C 20 | 205.8 | 21.2 | 364 | 10 | BF073878 |
| C 21 | 200.8 | 20.7 | 370 | 9 | AI716505 |
| C 22 | 200.4 | 20.6 | 3241 | 11 | AK035543 |
| C 23 | 199.4 | 20.5 | 485 | 14 | CB727068 |
| C 24 | 196.8 | 20.3 | 401 | 13 | BY438883 |
| C 25 | 192 | 19.8 | 359 | 12 | BM124337 |
| C 26 | 190.6 | 19.6 | 355 | 12 | BM123571 |
| C 27 | 188.4 | 19.4 | 345 | 10 | AW784908 |
| C 28 | 175.6 | 18.1 | 357 | 10 | BE667930 |
| C 29 | 164.4 | 16.9 | 621 | 12 | EG792248 |
| C 30 | 156.4 | 16.1 | 1030 | 12 | BM552361 |
| C 31 | 150 | 15.4 | 397 | 10 | AW479627 |
| C 32 | 147.2 | 15.2 | 276 | 9 | AI763601 |
| C 33 | 147 | 15.1 | 410 | 9 | AI716969 |
| C 34 | 144.2 | 14.9 | 569 | 10 | BE757743 |
| C 35 | 144 | 14.8 | 274 | 13 | EQ943127 |
| C 36 | 138.4 | 14.3 | 749 | 14 | CB494554 |
| C 37 | 138 | 14.2 | 304 | 12 | BM031895 |
| C 38 | 137.8 | 14.2 | 629 | 12 | BI429497 |
| C 39 | 135.8 | 14.0 | 553 | 12 | BM185681 |
| C 40 | 133.8 | 13.8 | 683 | 10 | AW862431 |
| C 41 | 128.4 | 13.2 | 995 | 13 | EQ722567 |
| C 42 | 128.2 | 13.2 | 187 | 10 | AW435707 |
| C 43 | 123 | 12.7 | 392 | 12 | BM186727 |
| C 44 | 122.4 | 12.6 | 750 | 28 | BZ173917 |
| C 45 | 120.8 | 12.4 | 567 | 12 | BI345924 |

ALIGNMENTS

RESULT 1

BM985339/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BM985339 715 bp mRNA linear EST 20-FEB-2003
UI-CF-EC1-acg-j-11-0-UI-s1 UI-CF-EC1 Homo sapiens cDNA clone
UI-CF-EC1-acg-j-11-0-UI-3', mRNA sequence.

BM985339

BM985339.1 GI:19611739

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 715)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

889546

Contact: McCray, PB

McCray Lab

University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 (www.openbiosystems.com).
 Seq primer: M13 FORWARD
 F0b1A=Yes.

FEATURES

source

1. 715
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-ECL-acg-j-11-0-UI"
 /tissue_type="lung"
 /dev_stage="Adult and Fetal"
 /lab_hosts="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-ECL"
 /notes="Organ: Lung; Vector: pT713-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-CF-ECL is a normalized cDNA library containing the
 following tissue(s): Normal lung from adult and from fetal
 day 64, day 87, week 19 and week 42. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pT713-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 AAGTGCCTTAC.
 TAG TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
 and 380-383
 TAG LIB=UI-CF-ECL
 TAG SEQ=AAGTGCCTTAC"

ORIGIN

Query Match 73.5%; Score 713.4; DB 12; Length 715;
 Best Local Similarity 99.9%; Pred. No. 1.3e-81;
 Matches 714; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 210 ACTCAGAAATCCGAGGTGAAGACTACTGTGGCAGATGTTGAGTGCACCCAGTACCTGC 269
 DB 715 ACTCAGAAATCCGAGGTGAAGACTACTGTGGCAGATGTTGAGTGCACCCAGTACCTGC 656

QY 270 ACAACACGACATCCTGCACCTGCAGCTCGAGTCCGAGAACATGATCATCCGGAATACA 329
 DB 655 ACAACACGACATCCTGCACCTGCAGCTCGAGTCCGAGAACATGATCATCCGGAATACA 596

QY 330 ACTTGCTCAAGGTGCTGGACCTCGGCAATGACAGAGCTTCAGCCAGGAGAAGTGTCTGC 389
 DB 595 ACCTGCTCAAGGTGCTGGACCTCGGCAATGACAGAGCTTCAGCCAGGAGAAGTGTCTGC 536

QY 390 CCTCAGACATGTCAGGACTACTAGACATGCTCGAGTCCGAGTCCCTGGAGGCGCAGG 449
 DB 535 CCTCAGACATGTCAGGACTACTAGACATGCTCGAGTCCGAGTCCCTGGAGGCGCAGG 476

QY 450 GGGCTGTTCACAGACAGACATCTGGGCCATCGGTGTGACAGCTTTCATCATGCTGAGCG 509
 DB 475 GGGCTGTTCACAGACAGACATCTGGGCCATCGGTGTGACAGCTTTCATCATGCTGAGCG 416

QY 510 CCGAGTACCCGCTGAGCAGGAGGTGTCACCGACCTGACAGAGAGACTGCCCAAGGGCG 569
 DB 415 CCGAGTACCCGCTGAGCAGGAGGTGTCACCGACCTGACAGAGAGACTGCCCAAGGGCG 356

QY 570 TGGTCCGGCTGAGCCGCTGTACCGGGGCTGTCCGGGGCGCGCTGGGCTTTCCTGCGCA 629
 DB 355 TGGTCCGGCTGAGCCGCTGTACCGGGGCTGTCCGGGGCGCGCTGGGCTTTCCTGCGCA 296

QY 630 GCATCTGTGCGGCCAGCCCTCGGGCGCGCTTCGCGCTGCAGCTGCCTGCAGTGCCTCG 689

Db 235 GCATCTGTGCGCCAGCCCTGGGGCGCGCCCTCGGGTCCAGCTGCTGAGTGCCTG 236
 QY 690 GGCTAACACAGAGAGAGGGCGCGCTGTTTGGCGGCCCGCGCGCTGACCTTCCCTACCGCGC 749
 Db 235 GGCTAACACAGAGAGAGGGCGCGCTGTTTGGCGGCCCGCGCGCTGACCTTCCCTACCGCGC 176
 QY 750 GGCTGCGCGCTTCTGTCGGCAATCGCAGAGAGAGAGCGCGCTGCTGTACAGAGGACACA 809
 Db 175 GGCTGCGCGCTTCTGTCGGCAATCGCAGAGAGAGAGCGCGCTGCTGTACAGAGGACACA 116
 QY 810 ACCTGGGCCAGGTGCGCTGAGGGTTCGCCCGGCCGACACCTTGGTCTCCCGCTGGGGGT 869
 Db 115 ACCTGGGCCAGGTGCGCTGAGGGTTCGCCCGGCCGACACCTTGGTCTCCCGCTGGGGGT 56
 QY 870 CGCTGCAGACCGCGCAATAAAGCGCCAGCGCGCGGCGAGAAAAA 924
 Db 55 CGCTGCAGACCGCGCAATAAAGCGCGCGGCGAGAAAAA 1

RESULT 2
 BG818748 731 bp mRNA linear EST 22-MAY-2001
 LOCUS 60279074F2 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4914402
 DEFINITION 5', mRNA sequence.
 ACCESSION BG818748
 VERSION BG818748
 KEYWORDS EST
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 731)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10821 row: d column: 19
 High quality sequence stop: 693.

FEATURES
 source
 1. 731
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4914402"
 /tissue_type="anaplastic oligodendroglioma with ip/19q
 loss"
 /lab_hosts="DH10B (T1 phage-resistant)"
 /clone_lib="NCI_CGAP_Brn67"
 /note="Organ: Brain; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.3 kb. Constructed by Life
 Technologies. Note: this is a NCI CGAP Library."

ORIGIN
 Query Match 69.2%; Score 672.2; DB 12; Length 731;
 Best Local Similarity 97.9%; Pred. No. 2.2e-76;
 Matches 702; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

QY 220 CGAGGTGAAGAGACTACCTGTGGCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGCA 279
 Db 1 CGAGGTGAAGAGACTACCTGTGGCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGCA 60

QY 280 CATCTCTGACCTGACCTGAGGTCCGAGACATGATCATCCGGAATACCAACCTGCTCA 339
 Db 61 CATCTCTGACCTGACCTGAGGTCCGAGACATGATCATCCGGAATACCAACCTGCTCA 120

```
QY 340 GGTCTGTGACCTGGGCAATGACAGAGCCTCAGCCAGAGAGAGGTGTGTCGCCCTCAGACAA 399
|
|
|
Db 121 GGTCTGTGACCTGGGCAATGACAGAGCCTCAGCCAGAGAGAGGTGTGTCGCCCTCAGACAA 180
|
|
|
QY 400 GTTCAAGACTACCTAGAGACCATGGCTCCAGAGCTCTGAGAGGCGAGGGGGCTGTTC 459
|
|
|
Db 181 GTTCAAGACTACCTAGAGACCATGGCTCCAGAGCTCTGAGAGGCGA-GGGGCTGTTC 239
|
|
|
QY 460 ACAGACAGACATCTGGGCCATCGGTGTGACAGCCTTCATCATGTGAGCGCCGAGTACCC 519
|
|
|
Db 240 ACAGACAGACATCTGGGCCATCGGTGTGACAGCCTTCATCATGTGAGCGCCGAGTACCC 299
|
|
|
QY 520 GGTGAGCAGCAGAGGTGCACGCGACCTGCAGAGGAGCTGGCAGGGGCTGTGTCGCCGT 579
|
|
|
Db 300 GGTGAGCAGCAGAGGTGCACGCGACCTGCAGAGGAGCTGGCAGGGGCTGTGTCGCCGT 359
|
|
|
QY 580 GAGCGCTCTACGCGGGGCTGTTCGGGGGGCGGCTGGCCCTTCCTGGCGACACTCTGTG 639
|
|
|
Db 360 GAGCGCTCTACGCGGGGCTGTTCGGGGGGCGGCTGGCCCTTCCTGGCGACACTCTGTG 419
|
|
|
QY 640 CGCCAGCCCTGGGCGCGGCTGTGCGGTTCAGCTGCTCAGTGGCGGTGAGTAAAGAGA 699
|
|
|
Db 420 CGCCAGCCCTGGGCGCGGCTGTGCGGTTCAGCTGCTCAGTGGCGGTGAGTAAAGAGA 478
|
|
|
QY 700 GGAGGGCCCGGCTGTTCGGGGCGGCGGCTGACCTTCCTACCGCGGCGGCTGCGCGT 759
|
|
|
Db 479 GGAGGGCCCGGCTGTTCGGGGCGGCGGCTGACCTTCCTACCGCGGCGGCTGCGCGT 538
|
|
|
QY 760 CTTCTGTGCGCAATCGCAGAGAGAGACGCGCTCTCTTACAGAGGACACACTGGGCCA 819
|
|
|
Db 539 CTTCTGTGCGCAATCGCAGAGAGAGACGCGCTCTCTTACAGAGGACACACTGGGCCA 598
|
|
|
QY 820 GGTGCGCTGAGGTGCGCCCGGCGACACCTTGTCTCTCCCGCTGGGGGTGCTGTCAGAC 879
|
|
|
Db 599 GGTGCGCTGAGGTGCGCCCGGCGACACCTTGTCTCTCCCGCTGGGGGTGCTGTCAGAC 658
|
|
|
QY 880 CGGCCAATAAAGCCAGCGCGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 936
|
|
|
Db 659 CGGCCAATAAAGCCAGCGCGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 715
|
|
|
```

```
RESULT 3
BX435097/c
LOCUS
DEFINITION
  BX435097 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
  CS0DF017YK06 3-PRIME, mRNA sequence.
ACCESSION
  BX435097
VERSION
  BX435097.1 GI:30783342
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 989)
  Li.W.B., Gruber.C., Jessee.J. and Polayes,D.
  Full-length cDNA libraries and normalization
  Unpublished (2001)
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 191 91006 EVRY cedex - France
  Email: segret@genoscope.cns.fr, Web : www.genoscope.cns.fr
  Library was constructed by Life Technologies, a division of
  Invitrogen. This sequence belongs to sequence cluster 1025.r For
  more information about this cluster, see
  http://www.genoscope.cns.fr/
  cgi-bin/cluster.cgi?seq=CS0BAK036AC04NM1&cluster=1025.r. Contact :
  Peng Liang Email : liang@lifetech.com URL : Corporation 1600
  http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
  Faraday Avenue Genoscope sequence ID : CS0BAK036AC04NM1.
  Location/Qualifiers
    1..989
    /organism="Homo sapiens"
```

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF017YK06"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
```

ORIGIN

```
Query Match      64.5%; Score 626.4; DB 13; Length 989;
Best Local Similarity 96.3%; Pred. No. 1.3e-70;
Matches 674; Conservative 0; Mismatches 21; Indels 5; Gaps 3;

QY 200 AGGCGCTCTCTACTCAGAA--TCCGAGGTGAAGGACTTACCTGTGGCAGAGATGTTGAGTGCCA 257
|
|
|
Db 755 AGGCGCTCTCAGCTCAGAAACTCGAGGTGAAGGGCTACCTGTGGCAGAGATGTTGAGTGCCA 696
|
|
|
QY 258 CCCAGTACCTGCACAAACACAGCAG-ATCCTGACACCTGGACCTGAGGTCCGAGAACATGATC 316
|
|
|
Db 695 CCCAGTACCTGCACAAACACAGCAGCAATCCTGCACCTGGACCTGAGGTCCGAGAACATGATC 636
|
|
|
QY 317 ATCACCAGATACAACTGCTCAAGGTCTGGAACCTGGGCAATGACAGAGCCCTCAGCCAG 376
|
|
|
Db 635 ATCACCAGATTCACTGCTCAAGGTCTGGAACCTGGGCAATGACAGAGCCCTCAGCCAG 576
|
|
|
QY 377 GAGAAAGTCTGCTCCCTCAGACCAAGTTCAAGACTACCTAGAGACCAATGCTCCAGAGCTC 436
|
|
|
Db 575 GAGAAAGTCTGCTCCCTCAGACCAAGTTCAAGACTACCTAGAGACCAATGCTCCAGAGCTC 516
|
|
|
QY 437 CTGAGGGGCGAGGGGCTGTTCCACAGACAGACATCTGGGCGCATCGGTGTGACAGCCTTC 496
|
|
|
Db 515 CTGAGGGGCGAGGGGCTGTTCCACAGACAGACATCTGGGCGCATCGGTGTGACAGCCTTC 456
|
|
|
QY 497 ATCATGCTGAGCGCCGAGTACCCGGTGAAGCGAGGGTGCACGCGACCTGACAGAGGA 556
|
|
|
Db 455 ATCATGCTGAGCGCGAGTACCCGGTGAAGCGAGGGTGCACGCGACCTGACAGAGGA 396
|
|
|
QY 557 CTGCGCAAGGGGCTGTCGGCTGAGCCCTGCTACCGGGGGGCTGTCCGGGGGCGCGCTG 616
|
|
|
Db 395 CTGCGCAAGGGGCTGTCGGCTGAGCCCTGCTACCGGGGGGCTGTCCGGGGGCGCGCTG 336
|
|
|
QY 617 GCCTTCTCGCGAGCAGCTCTGTGCGGCCAGCCCTGGGGCGGCGCTGCGCGCTCCAGCTGC 676
|
|
|
Db 335 GCCTTCTCGCGAGCAGCTCTGTGCGGCCAGCCCTGGGGCGGCGCTGCGCGCTCCAGCTGC 276
|
|
|
QY 677 CTGCACTGCCCGTGGTAAACAGAGAGGCGCCGGCTGTTCGCGCGCCCGCGCGCTGAC 736
|
|
|
Db 275 CTGCACTGCCCGTGGTAAACAGAGAGGCGCCGGCTGTTCGCGCGCCCGCGCGCTGAC 216
|
|
|
QY 737 TTCCCTTACCGCGCGCTGCGCTCTTCGTGCGCAATCGCGAGAGAGACGCGCGCTGCTG 796
|
|
|
Db 215 TTCCCTTACCGCGCGCTGCGCTCTTCGTGCGCAATCGCGAGAGAGACGCGCGCTGCTG 156
|
|
|
QY 797 TACAAGAGGCACAACTGGCCAGAGTGCCTGAGGCTCGCCCGGCGCACACCTTGGTCT 856
|
|
|
Db 155 TACAAGAGGCACAACTGGCCAGAGTGCCTGAGGCTCGCCCGGCGCACACCT--GGTC 98
|
|
|
QY 857 CCCCCTGCGGGTCTGCTGACAGCGCGCCCAATAAAGCGC 896
|
|
|
Db 97 TCCCCTGCGGGTCTGCTGACAGCGCGCCCAAAAAAAGGGC 58
|
|
|
```

```
RESULT 4
BX280322
LOCUS
DEFINITION
  BX280322 NC1 CGAP_Brn67 Homo sapiens cDNA clone IMAGE5914402; mRNA sequence.
ACCESSION
  BX280322
VERSION
  BX280322.1 GI:28612364
```

```

BX280322 480 bp mRNA linear EST 04-MAR-2003
BX280322 NC1 CGAP_Brn67 Homo sapiens cDNA clone IMAGE5914402; mRNA sequence.
IMAGE:4914402, mRNA sequence.
BX280322
BX280322.1 GI:28612364
```

KEYWORDS: EST.
SOURCE: Homo sapiens (human)
ORGANISM: Homo sapiens
REFERENCE: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS: 1 (bases 1 to 480)
Ebert L., Heil O., Hennig S., Neubert P., Partsch E., Peters M., Radelof U., Schneider D. and Korn B.
TITLE: Human Unigeneset - RZPD3
JOURNAL: Unpublished (2003)
COMMENT: RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD: IMAGp998D1910821.
RZPDLIB: I.M.A.G.E. cDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-bin/showlib.pl.cgi?response?libNo=972 Contact: Ina Rofls
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13u, Primer sequence: CGTTGTAACACGACGCCAGT.

FEATURES
source
1..480
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGp998D1910821"; IMAGE:4914402"
/tissue_type="anaplastic oligodendroglioma with ip/19q loss"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Brn67"
/note="Organ: Brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."

ORIGIN
Query Match 49.4%; Score 480; DB 13; Length 480;
Best Local Similarity 100.0%; Pred. No. 6.6e-52;
Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
219 CCGAGGTGAAGGACTACTGTGGCAGATGTTGAGTGCCACCCAGTACTCGACCAACGAGC 278
1 CCGAGGTGAAGGACTACTGTGGCAGATGTTGAGTGCCACCCAGTACTCGACCAACGAGC 60
279 ACATCTGCACTGGACCTGAGTCCGAGAACATCATCATCCAGATACCATGCTGCTCA 338
61 ACATCTGCACTGGACCTGAGTCCGAGAACATCATCATCCAGATACCATGCTGCTCA 120
339 AGGTCTGGACCTGGGCAATCCAGAGACCTCAGCAGGAGAGGTGCTGCCCTCAGACA 398
121 AGGTCTGGACCTGGGCAATCCAGAGACCTCAGCAGGAGAGGTGCTGCCCTCAGACA 180
399 AGTTCAGAGGACTACTAGAGACCATGCTCCAGAGCTCTCTGGAGGGCCAGGGGCTGTTT 458
181 AGTTCAGAGGACTACTAGAGACCATGCTCCAGAGCTCTCTGGAGGGCCAGGGGCTGTTT 240
459 CACAGACAGACATCTGGGCCATCGGTGTGACAGCTTTCATCATGCTGAGGCGCCAGTACC 518
241 CACAGACAGACATCTGGGCCATCGGTGTGACAGCTTTCATCATGCTGAGGCGCCAGTACC 300
519 CGGTGAGCAGCGAGGGTGCACCGACCTGCGAGAGAGACTGCGCAAGGGCTGGTCCGGC 578
301 CGGTGAGCAGCGAGGGTGCACCGACCTGCGAGAGAGACTGCGCAAGGGCTGGTCCGGC 360
579 TGAGCCGCTGCTACGGGGCTGTCCGGGGCGCCGTGGCTTCTCTCGCAGACACTGTGT 638
361 TGAGCCGCTGCTACGGGGCTGTCCGGGGCGCCGTGGCTTCTCTCGCAGACACTGTGT 420

QY 639 GGGCCAGCCCTGGGCGGCGCCCTGCGGCTCCAGCTGCTGAGTGGCCGCTGCTAAG 698
Db 421 GGGCCAGCCCTGGGCGGCGGCGCCCTGCGGCTCCAGCTGCTGAGTGGCCGCTGCTAAG 480

RESULT 5
CD774776/c
LOCUS: CD774776 640 bp mRNA linear EST 02-JUL-2003
DEFINITION: UI-M-AQ0-cit-f-24-0-UI.s1 NIH_BMAP_MHI Mus musculus CDNA Clone
UI-M-AQ0-cit-f-24-0-UI 3', mRNA sequence.
ACCESSION: CD774776
VERSION: CD774776.1 GI:32433278
KEYWORDS: EST.
SOURCE: Mus musculus (house mouse)
ORGANISM: Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 640)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE: Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL: Genome Res. 6 (9), 791-806 (1996)
MEDLINE: 97044477
PUBMED: 8889548
COMMENT: Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20852-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
1..640
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AQ0-cit-f-24-0-UI"
/dev_stages="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_MHI"
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_MHI library is a non-normalized library constructed from mouse hippocampus. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.
TAG_TISSUE=hippocampus
TAG_LIB=UI-M-AQ0
TAG_SEQ=TTCCA"

ORIGIN
Query Match 41.2%; Score 399.8; DB 14; Length 640;
Best Local Similarity 77.4%; Pred. No. 9.5e-42;
Matches 498; Conservative 0; Mismatches 142; Indels 3; Gaps 1;
281 ATCTCTGCACTGGACCTGAGTCCGAGAACATGATCATCCGAATACACTGCTCAAG 340
640 ATCTCTGCACTGGACCTGAGTCCGAGAACATGATGCTCACTGAGTACCACTGCTTAAG 581
341 GTCGTGGACCTGGGCAATGCACAGACCTCAGCAGGAGAGGTGCTGCCCTCAGACAAG 400

Db 580 GTTATAGACCTGGGAAATGCCAGAGTCTCGACAGAGAGAGTCCCGACGCCCTTGAGAAC 521
QY 401 TTCAGAGGACTACCTAGAGACCAATGGCTCCAGAGTCTCTGGAGGGCGAGGGGCTGTTC 460
Db 520 TTCAGAGGACTACCTGGAGACCAATGGCTCCAGAGTCTCTGGAGGGCGAGGGGCTGTTC 461
QY 461 CAGACAGACATCTGGGCCATCGGTGTGACAGCCTTCATCAGCCTTCATCAGCCTGAGCGCCAGTACCG 520
Db 460 CAGACAGACATCTGGGCTATTGGTGTACAGCCTTCATTAATGCTGAGTGGGAGTACCA 401
QY 521 GTGACGAGGAGGTGACAGCACTGTGAGAGAGACTGTGCGAAGGGGCTGTGCGGCTG 580
Db 400 GAGAGCAGCGAGGGGACTTCGCACTGTGAGAAAGGCTTGGCGAAGGACTCATTCGGTTG 341
QY 581 AGCCGCTGTACGCGGGGCTGTCCGGGGCGCGGTGGGCTTCTTCCGCGAGCAGCTGTGC 640
Db 340 AGTCGCTGTATGAGGATTATCAGAGAGAGCGGTAGCTTCTTCCAGAGTTCATTAAT 281
QY 641 GCCAGGCTGTGGGCGGCTGTGCGGCTGTCAGTCTGTCAGTCTGCGGCTGCTAACAGAG 700
Db 280 GCTCAACCTGTGGGCGGCTGTGCGGCTTCCACCTGTTCGAATGGGATGGCTGACAGAG 221
QY 701 GAGGCGCGGCTGTTCGCGGCGCGGCTGTGAGTCTTCCAGTCTTCCAGTCTGCGGCTG 760
Db 220 GAGGCGCGGCTGTTCGCGGCGCGGCTGTGAGTCTTCCAGTCTTCCAGTCTGCGGCTG 161
QY 761 TTCGTGCGCAATCGCGAGAGAGAGCGCGCTGTGTACAGAGAGGACCAACTCGGCCAG 820
Db 160 TTGTGCGCGAGCGAGAGAGCGCGCTGTGTACAGAGAGGACCAACTCGGCCAG 101
QY 821 GTGCGCTAGGCTGCGGCGGCTGTGTGTGCTGCTGCGGCTGCGGCTGCGGCTGCGG 880
Db 100 GTGCGCTAGG---TCCAGCTCTACAGAGCAAGATGTGCCATGCCATGCGGGGACACC 44
QY 881 CGCAATAAAGCGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 923
Db 43 CGCTAATAAGAGTCAAGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1

RESULT 6
CF176441/c 622 bp mRNA linear EST 28-JUL-2003
LOCUS 800530 MARC 3P1G Sus scrofa cDNA 3', mRNA sequence.
DEFINITION CF176441
VERSION CF176441.1 GI:33288217
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 622)
Smith,R.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,
Wise,T.A., Noneman,D.J., Wray,J.E. and Keele,J.W.
A second set of porcine ESTs from a pooled-tissue normalized
library

JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: SRG8004 row: M column: 6
Seq primer: TAGAAGGCACAGTCGAGG.
Location/Qualifiers
1 .622
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"

FEATURES
source
1 .622
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"

/lab_host="DH10B"
/clone_lib="MARC 3P1G"
/notes="Vector: pcDNA3.1; Site: 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."
ORIGIN
Query Match 38.9%; Score 377.4; DB 14; Length 622;
Best Local Similarity 81.9%; Pred. No. 6.8e-39;
Matches 435; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 299 AGTCCGAGAAATGATCATCACCGAATACACCTCTCAAGGTCTGAGCACTGGGCAAT 358
Db 621 AGTCCGAGAAATGATCATCACCGAATACACCTCTCAAGGTCTGAGCACTGGGCAAT 562
QY 359 GCACAGAGCTCAGCAGAGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAG 418
Db 561 GCCCGAGATTTCATCTCAGGAGAGTGTCTGCCCTCAGACGGTTCAAGGACTATGTGGAG 502
QY 419 ACCATGGCTCCAGAGCTCTGGAGGCGCAGGGGCTGTTCACACAGACACATCTGGGCC 478
Db 501 ACTATGGCTCCAGAGCTCATAGAGGCGCAGGGTGTCTCCACAGACTGATATATGGGCC 442
QY 479 ATCGGTGTGACAGCTTCATCATGTAGCGCGGAGTACCCGGTGTGACGAGGAGGTGCA 538
Db 441 ATAGGTGTCACTGCCCTTCATCATGTAGTGTAGTACCCAGTGTGAGCGAGGGGACG 382
QY 539 CGCGACTGTGACAGAGAGTGTGCGAAGGGGCTGTCCGCTGAGCGCTGCTACGCGGG 598
Db 381 AGAGATTTCAGAGAAAGGCTTGGCGAAGGGGCTCATCCAGATGAGCGGATGCTACGCGGG 322
QY 599 CTGTCCGGGGGGCGGCTGTGCCCTTCTCTGCGCAGCACTCTGTGCGGCCAGGCCCTGGGGCGCG 658
Db 321 CTCTCCGGGGGTGCGCTGTGCCCTTCTCTGCGCAGCACTCTGTGCGGCCAGGCCCTGGGGCGCG 262
QY 659 CCCTGCGCTGTGAGTGTGCGCTGTGCGCTGTGCGCTGTGCGCTGTGCGCTGTGCGCTGTGCG 718
Db 261 CCATGTGCTGCAACTGTCTGCACTGCTGCACTGCTGCACTGCTGCACTGCTGCACTGCTGCA 202
QY 719 CGGCGCGCGCGCTGTGCTTCTCTGCGCGCGCTGTGCGCGCGCTGTGCGCGCGCTGTGCGCGCG 778
Db 201 CACCCAGCACTGTGACCTTCCACAGCAGCACTGTGCGCGCGCTGTGCGCGCGCTGTGCGCGCG 142
QY 779 AAGAGAGCGCGCTGTGTACAGAGGACAACTGTGCGCGCGCGCTGTGCGCGCGCTGTGCGCGCG 829
Db 141 AAGAGGCGGCGCTGTGTACAAAGGACAACTGTGCGCGCGCGCTGTGCGCGCGCTGTGCGCGCG 91

RESULT 7
BF134040 768 bp mRNA linear EST 24-OCT-2000
LOCUS 601778492F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:406702 5',
DEFINITION mRNA sequence.
ACCESSION BF134040
VERSION BF134040.1 GI:10973080
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 768)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
 Plate: LIA9238 row: 9 column: 23
 High quality sequence stop: 736.

FEATURES
 source
 1. 768
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="Czech 11"
 /db_xref="taxon:10090"
 /clone="IMAGE:4006702"
 /tissue_type="tumor, metastatic to mammary"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Lu30"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; transgenic model MMTV-1, expression driven by MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies."
 ; Investigator providing samples: Gilbert Smith, NIH"

ORIGIN
 Query Match 34.9%; Score 339; DB 10; Length 768;
 Best Local Similarity 79.6%; Pred. No. 4.8e-34;
 Matches 450; Conservative 0; Mismatches 110; Indels 5; Gaps 4;

QY 2 AGCGGGGGGGCTGGCGCCCAAGCATCATCCCTTACCACCCCAAGGACAGACAGCAGTG 61
 DB 179 AGTGGCGGGCCCTAGCTGCTAAGATCGTTCCTTACCACCTGAGGACAGACAGCTGTA 238
 QY 62 CTGCGCGAATACGAGGCGCTCAAGGGCTGGCGCCACCCGACCTGGCCCGACCTGCGCGCA 121
 DB 239 CTAAGAGAATACGAGGCATTAAGAGATGACACCCACATCTGGCCCAACTCCATGCC 298
 QY 122 GCCTACCTCAGCCCGCCGACCTGGTCTCATCTTGGAGCTGTCTCTGGGCCCGAGCTG 181
 DB 299 GCCTACCTCAGTCCCGGCACCTGGTCTCATCTGGAGCTGTCTCTGGCCCTGAGCTG 358
 QY 182 CTCCTCTGCTGGCGGAGGGCTCTCTACTCAGAAATCCGAGGTGAAGACTACCTGTGG 241
 DB 359 CTACCTCTTTGGGGGAGGAGGAATCTTACTCAGAGTCTGATGTGAAGGACTACCTGTGG 418
 QY 242 CAGATGTTGAGTCCACCCAGTACTCTGCACACAGACACATCTTGGACCTGGACCTGAGG 301
 DB 419 CAGATGTTGAGTCCACCCAGTACTCTGCACACAGACACATCTTGGACCTGGACCTGAGG 478
 QY 302 TCCGAGAACATGATCATCACCGAATACAACCTGCTCAAGTCTGTGGACCTGGCGCAATGCA 361
 DB 479 TCCGAGAACATGATGTTCACTGAGTACAACCTGCTTAAGTTATAGACCTGGGAATGCC 538
 QY 362 CAGAGCTTCAGCAGGAGAGGTGCTGCTCTCAGACAGATTCAGAGGACTACCTAGAGACC 421
 DB 539 CAGAGTCTCGACCAAGAGAAGGTCCAGCCCTCAGAGACTTCAAGAGACTACCTGGAGACC 598
 QY 422 ATGGCTCCAGAGCTCTCTGGAGGGCCAGGGGGCTGTTCACAGACAGACATCTGGGCCATC 481
 DB 599 ATGGCTCCAGAACTCTTGAAGGTTCACCGGCTGTTCACAGACAGACA--TTGGGCTAT 656
 QY 482 GGTGTGACAGCCTTATCATCTGAGCGCCGAGTACCGGTGAGCAGGAGGGTGCACGC 541
 DB 657 GGTGTAAACAG-CTTCTATTATCTGAGTGGGAGTA-CCAGAGAGCAGGAGGGGAGCTCGT 714
 QY 542 GACCTGCAGAGAGGAGTCCGCAAGG 566
 DB 715 GA-CTGGAGAAAGGCTTCCGAGGG 738

RESULT 8
 BF651426
 LOCUS BF651426 479 bp mRNA linear EST 25-APR-2001
 DEFINITION 274209 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BF651426
 VERSION BF651426.1 GI:11916556
 KEYWORDS EST.
 SOURCE Bos taurus (cow)

ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 Bovidae; Bovinae; Bos.

REFERENCE
 1 (bases 1 to 479)
 Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
 Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C.,
 Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A.,
 Chitko-McKown, C.G., Pettea, G., Holt, I., Karamycheva, S., Liang, F.,
 Quackenbush, J., and Keele, J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT

Contact: Smith RPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithr@mail.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACGTCACGACG
 Plate: 63 row: L column: 10
 Seq primer: ATTAGTGACACTATAG.
 Location/Qualifiers
 1..479
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 3BOV"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from marrow, alveolar
 macrophage, ovary, fetal semitendinosus muscle, and fetal
 longissimus muscle."

ORIGIN
 Query Match 34.1%; Score 331; DB 10; Length 479;
 Best Local Similarity 82.0%; Pred. No. 5.8e-33;
 Matches 393; Conservative 0; Mismatches 85; Indels 1; Gaps 1;

QY 323 GAATACAACTGCTCAAGTCTGTGGACCTGGGCAATGCACAGAGCTCAGCCAGGAGAG 382
 DB 2 GAGTATACCTGCTC-AGTTCATCGACTTCGGAACGCCAGAGTTTGGCCAGGAGAGG 60
 QY 383 GTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGACCATGGCTCCAGAGCTCTGGAG 442
 DB 61 GTCTCTGCCCTCAGAGAGATTCAAGGACTACATGGAGACCATGGCTCCGAGCTCTGGAG 120
 QY 443 GSCCAGGGGGCTGTCCACAGACACATCTGGGCTCGGTGTGACAGCTTCATCATG 502
 DB 121 GSCCAGGGGGCTGTCTGGCAGACTGACATCTGGGCCATAGGTGTACAGCTTCATCATG 180
 QY 503 CTGAGCGCCGAGTACCCCGTGAAGCAGCGAGGCTGACCGGACCTGTGACAGAGAGACTGGCC 562
 DB 181 TTGAGCGCTGACTACCCCGTGAACCGTGAAGGAAACGAGAGACACGAGAGGCGCTGGC 240
 QY 563 AAGGGGCTGCTCCGCTGAGCGGCTGCTACGCGGGGCTGTCCGGGGGCGCGTGGCCCTTC 622
 DB 241 AAGGGGCTCATCCAGCTGAGCGGCTGCTATGCGGGGCTGTCTGGGGGCGCTGTGGCCCTTC 300
 QY 623 CTGCGCAGCACTCTGTGGCGCCAGCCCTGTGGGCGCGGCCCTGGCGGTCCAGTGTGCTGAG 682
 DB 301 CTCCGAAGCACAACCTGTGTGTGCACCCCTGGTCCGGCGGTGTGCATCCAGCTGCTGCGAG 360
 QY 683 TGCCTGTGCTAACAGAGGAGGCGCCGCTTCTCGCGGCGCGCGCTTCCCTTCCCT 742

```

Db      361 AGCCCGTGGTGAACGAGAGAGGGCCAGCCAGCTCCAGCCCGGGGGGTGACCTTCCCG 420
QY      743 ACCGCGCGCTGCGCGTTCCTGCGCAATCCGAGAGAGACGCGCGCTGCTGTACAA 801
Db      421 ACCACTGCTGCGCGTCTTTGTCGAGAGCGCGAAGAGAGCGCGGCACTGCTGTACAA 479

RESULT 9
BQ554402/c
LOCUS   H4027H08-3 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
DEFINITION
ACCESSION BQ554402
VERSION   BQ554402.1 GI:21455290
KEYWORDS EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE
AUTHORS   VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G.,
          Martin,P.R., Stagg,C.A., Bassey,U., Aiba,K., Hamatani,T.,
          Kargul,G.J., Luo,A.G., Keiso,J., Hide,W. and Ko,M.S.H.
          Assembly, verification, and initial annotation of NIA 7.4K mouse
          cDNA clone set
          Genome Res. 12 (12), 1999-2003 (2002)
          2354164
          12466305
          Other ESTs: H4027H08-5
          Contact: Yong Qian
          Laboratory of Genetics
          National Institute on Aging/National Institutes of Health
          333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
          Email: cdna@igsun.grc.nia.nih.gov
          This clone set has been freely distributed to the community. Please
          visit http://lgsun.grc.nia.nih.gov/cDNA/NIA\_7.4K.html for details.
          Plate: H4027 row: H column: 08
          Seq primer: -21M13 Forward
          High quality sequence stop: 522
          POLYA=Yes
FEATURES             Location/Qualifiers
     source           1..522
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="niaEST:H4027H08-3"
                     /db_xref="taxon:10090"
                     /clone="H4027H08"
                     /sex="mixed"
                     /dev_stage="mixed"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse 7.4K cDNA Clone Set"
                     /note="Vector: pSPORT1; Site_1: Sali; Site_2: NotI; This
clone is among a rearrayed set of 7,407 clones from more
than 20 cDNA libraries."
ORIGIN
Query Match      32.3%; Score 313.4; DB 13; Length 522;
Best Local Similarity 80.9%; Pred. No. 9,7e-31;
Matches 365; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY      381 AGGTGTCCTCCAGACAGTTCAAGACTACCTAGAGACATGGCTCCAGAGCTCTCTGG 440
Db      522 AGGTCCCGACGCCCTGAGAACTTCAAGACTACCTGGAGACCATGGCTCCAGAACTCTCTGG 463
QY      441 AGGGCCAGGGGGTGTCCACAGACACACATCTGGGCCATCGGTGTGACAGCTTCATCA 500
Db      462 AAGTCAAGGGGCTGTCCACAGACACACATTTGGCTATTTGGTGTAAAGCTTTCATTA 403
QY      501 TGCTGAGCGCGAGTACCCGGTGAAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 560
Db      402 TGCTGAGTGGCGAGTACCCAGAGAGCAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 343

```

```

QY      561 GCAAGGGGTGGTCCGGCTGAGCCGCTGTACGCGGGGGTGTCCGGGGGCGCGTGGCTTCC 620
Db      342 GCAAGGGACTCATTCGTTGAGTGTGCTATATGAGGATTAACAGAGGAGCGGTAGCCT 283
QY      621 TCTTGGCGAGCACTGTGTGCGCCAGCCCTGCGGCCCGCCCTGCGGGTCCAGCTGCTGCG 680
Db      282 TCTTGCAAGATTCAATATGTCTCAACCTGGGGGCGCGCTCCACCTGCTTGC 223
QY      681 AGTGCCCGGTGGTAAACAGAGGAGGGCCCGCTGTTCGCGGGCCCGCGCTGACCTTCC 740
Db      222 AATGCGGATGGTGCAGAGAGGAGGGCCCAACCGCTCCCGCCACGCGGTACCTTCC 163
QY      741 CTACCGCGGGGTGCGCGCTCTTCGTGCGCAATCGGAGAGAGAGCGCGGCTGTGTACA 800
Db      162 CCACCGTGGGTGCGCGCTTTTGTGCGGAGCGCGAGAAACGGGCTGCGCTCTCTACA 103
QY      801 AGAGGCACACCACTGGCCCGAGGTGCGCTGAGG 831
Db      102 AGAGGATTAACCTGGCCAGGTGCGCTGAGG 72

RESULT 10
BQ110731/c
LOCUS   BE110731-1 GI:8502877
DEFINITION
ACCESSION BE110731
VERSION   BE110731.1
KEYWORDS EST.
SOURCE    Rattus norvegicus (Norway rat)
ORGANISM  Rattus norvegicus
REFERENCE
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
          Normalization and subtraction: two approaches to facilitate gene
          discovery
          Genome Res. 6 (9), 791-806 (1996)
          97044477
          8889548
          Contact: Soares, MB
          Coordinated Laboratory for Computational Genomics
          University of Iowa
          375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
          Tel: 319 335 8250
          Fax: 319 335 9565
          Email: bento-soares@uiowa.edu
          The sequence contained an oligo-dT track that was present in the
          oligonucleotide that was used to prime the synthesis of first
          strand cDNA and therefore this may represent a bonafide poly A
          tail. The sequence tag present in the cDNA between the NotI site
          and the oligo-dT track served to identify it as a clone from the
          normalized heart library cDNA library Preparation: M.B. Soares Lab
          Clone distribution: clones will be available through Research
          Genetics (www.resgen.com)
          Seq primer: M13 Forward
          POLYA=Yes.
FEATURES             Location/Qualifiers
     source           1..493
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BJ1-avd-g-08-0-UI"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-BJ1"
                     /note="Vector: pYT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
library is a subtracted library derived from the following
tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.

```

For a detailed description of the library from which this clone was derived, please visit our web site at rategest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG TISSUE=heart
 TAG LIB=UI-R-BJ1
 TAG_SEQ=ACAAC"

ORIGIN

Query Match 32.1%; Score 311.4; DB 10; Length 493;
 Best Local Similarity 82.4%; Pred. No. 1.8e-30;
 Matches 357; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 407 GACTACCTAGAGACCATGGCTCCAGAGCTCCTGGAGGGCCAGGGGGTGTTCACAGACA 466
 Db |||||
 QY 493 GACTACCTAGAGACCATGGCTCCAGACTTCTGGAGGCCAGGGGGTTCACAGACA 434
 Db |||||
 QY 467 GACATCTGGGCGATCGGTGTGACAGCCCTTCATCATGTGAGCGCCGAGTACCCGGTGAGC 526
 Db |||||
 QY 433 GACATCTGGGCGATCGGTGTGACAGCCCTTCATCATGTGAGTGGCGAGTACCCAGTGAGC 374
 Db |||||
 QY 527 AGCAGAGGTGCACGCGACCTGCGAGAGGACTGCGAAGGGGTGTCGGGTGAGCCGC 586
 Db |||||
 QY 373 AGCAGAGGAGCTCGGACCTGCGAAGAGGCTGCGCAGGAGTCAATTCAGTGC 314
 Db |||||
 QY 587 TGCTACCGGGGGTGTCCGGGGCGCGTGGCCCTTCCTGCGAGCACTCTGTGCGCCAG 646
 Db |||||
 QY 313 TGCTATGAGGATATCAGGGGGTGGGTAGACCTTCCTGCGAGATTCATTGCGGTGCG 254
 Db |||||
 QY 647 CCTTGGGGCGGCGCTGCGGTGCGAGTGCCTGAGTCCCGTGGCTTACAGAGAGGGGC 706
 Db |||||
 QY 253 CCTTGGGGTGCCTGCGGTTCACCTGCTGAGTGGGGTGGCTGACGAGAGGGGC 194
 Db |||||
 QY 707 CCGGCGCTTTCGGGGCGGCGCGTGCCTTCCTACCGCGGGGTGCGCGTCTTCGTG 766
 Db |||||
 QY 193 CCGGCGCTTTCGGGGCGGCGCGTGCCTTCCTACCGCGGGGTGCGCGTCTTCGTG 134
 Db |||||
 QY 767 CGCAATCGGAGAGAGACGGCGCTCTGTACAGAGGACAACTGGCCAGGTGCGC 826
 Db |||||
 QY 133 CGCAGGCGGAGAGCGCGGGCGCTACTCTACAGAGACAACTGGCTCAGGTGCGC 74
 Db |||||
 QY 827 TGAGGGTGCGCC 839
 Db |||||
 QY 73 TGAGGGCGAGGCC 61
 Db |||||

RESULT 11
 CF177813
 LOCUS 494 bp mRNA linear EST 28-JUL-2003
 DEFINITION 805332 MARC 3P1G Sus scrofa cDNA 5', mRNA sequence.
 ACCESSION CF177813
 VERSION CF177813.1 GI:33289589
 KEYWORDS EST.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 494)
 Smith, T.P.L., Freking, B.A., Ford, J.J., Vallet, J.L., Fox, J., Wise, T.A., Nonneman, D.J., Wray, J.E. and Keele, J.W.
 A second set of porcine ESTs from a pooled-tissue normalized library
 Unpublished (2003)
 Contact: Smith, TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@meat.mars.usda.gov
 Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross_match v0.990329.

Plate: SR8004 row: M column: 6
 Seq primer: GTAATAGCACTCACTATAGG.

FEATURES

source

1..494
 Location/Qualifiers
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /db_xref="taxon:9823"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 3P1G"
 /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
 Library made with RNA pooled from multiple tissues
 including brain, liver, muscle, placenta/endometrium,
 ovary, testes, and bone marrow."

ORIGIN

Query Match 25.5%; Score 247.4; DB 14; Length 494;
 Best Local Similarity 82.3%; Pred. No. 2.4e-22;
 Matches 284; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 1 CAGCGGCGGCGCTGGCGCCCAAGATCATCCCTACACCCCAAGACAGACAGT 60
 Db |||||
 QY 150 CATTGGGCGCGTGTGGCTGCAATATCGTGGGGGCGCCCTGAGAACAGGACTGCCGT 209
 Db |||||
 QY 61 GCTGCGGCAATACGAGGCGCTCAAGGGCTGCGCCACCCGACCTGCCAGCTGCACGC 120
 Db |||||
 QY 210 GCTTCGATAATATGATGCTTCAAGGGCTTGGCCACCCTCACCTGGCACAATTGCGAGC 269
 Db |||||
 QY 121 AGCTTACCTCAGCCCGCGCACCTGGTGTCTCATCTTGAGGTGTCTCTGGGCCGAGCT 180
 Db |||||
 QY 270 TGCCTACCTCATCCCGGCGACCTGGTCTCATCTTGAGTGTGTCTGGCGCTGATCT 329
 Db |||||
 QY 181 GCTCCCTGCTGCGCGAGGGGCTCTACTAGATCCGAGTCCGAGTGAAGCACTACCTGTG 240
 Db |||||
 QY 330 GCTCCCTGCTGCGCGAGGGGCTCTACTATATATATATATATATATATATATATATAT 389
 Db |||||
 QY 241 GCAGATGTTGAGTGCACCCAGTACCTGCACCAACAGCACATCTGCACCTGACCTGAG 300
 Db |||||
 QY 390 GCAGATGCTGATGGCACTATTACCTGCAGCCGCGATCTGCTGATCTAGACCTCAT 449
 Db |||||
 QY 301 GTCGAGAACATGATCATCCGCAATACACCTGCTCAAGTCT 345
 Db |||||
 QY 450 GTCCGAAAACATGATGTTGCACCGAGTACACCTGCTCAAGATCAT 494
 Db |||||

RESULT 12

AI503993/c

LOCUS

428 bp mRNA linear EST 11-MAR-1999

DEFINITION

vm45d02.x1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
 clone IMAGE:1001187 3', mRNA sequence.

ACCESSION

AI503993

VERSION

AI503993.1 GI:4401844

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 428)

REFERENCE

1 (bases 1 to 428)
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swales, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, I., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)

JOURNAL

Contact: Marta M/WashU-NCI Mouse EST Project 1999

COMMENT

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:565403
This clone was previously sequenced on the 5' end only, this new data is from the 3' end
High quality sequence stop: 427.

FEATURES

source
1..428
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:1001187"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse diaphragm (#937303)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA prepared from diaphragm muscle. Primer: Oligo dt. Average insert size: 1.5 Kb. Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 25.1%; Score 243.4; DB 9; Length 428;
Best Local Similarity 80.1%; Pred. No. 8.1e-22;
Matches 286; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 475 GCCATCGGTGACAGCCTTCATCATGCTGAGCGCCAGTACCGGTGACGAGGG 534
DB 428 GGCATTGGTGTAAAGAGCTTCATTATGCTGAGTGGCAGTACCCAGAGACAGGG 369
QY 535 TGCAGCGACCTGACAGAGACTGCGAAGGGGTGCTCGGTGACCGCTGACGC 594
DB 368 GACTCGCACCTGCAGAAAGGTTGCGCAAGGACTATTGGTTAGTCGCTATGC 309
QY 595 GGGGCTGTCGGGGCGCGCTTCTCGCGAGCACTGTGCGCCAGCCCTGGG 654
DB 308 AGGATTATCAGAGAGCGGTAGCCCTCTCGAGAGTTCATTATGTCTCAACCTGGG 249
QY 655 CGGCCCTCGCGTCCAGCTGCTGAGTGGCGGTGCTAAACAGAGAGGGCCGGCTG 714
DB 248 CGGCCGTGGCTTCCACCTGCTGCAATGCGGTGCTGACAGAGAGGGCCACCG 189
QY 715 TTGCGGGCGCGCGCGTTCCTACCGCGCGGTGCGGTCTTCGTGCGCAATCG 774
DB 188 CTCGCGGCCAGCGCCGTAACCTTCCCGCGGTGCGGTGCGCGCTTTGTGCGCAGCG 129
QY 775 CGAGAAGAGAGCGCGCTGTGTACAAGAGGCACAACTGCGCCAGGTGCGCTGAGG 831
DB 128 CGAGAAGCGCGTGGCTCTCTCTACAGAAGCATATCTGCCCCAGGTGCGCTGAGG 72

RESULT 13
BF414891/c
LOCUS
DEFINITION
UI-R-BJ2-bot-c-07-0-UI-s1 UI-R-BJ2 Rattus norvegicus cDNA clone
ACCESSION
BF414891
VERSION
BF414891.1 GI:11402880
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 408)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
9704477
PUBMED
889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics

University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized heart library cDNA library Preparation: M.B. Soares Lab
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source
1..408
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ2-bot-c-07-0-UI"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BJ2"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ2 library is a subtracted library derived from the following tissues: heart, atrium at 15 dpc, ventricle at 15.5 dpc, atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15 dpc, AV canal at 15 dpc. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_TISSUE=heart
TAG_LIB=UI-R-BJ2
TAG_SEQ=ACAAC"

ORIGIN

Query Match 24.5%; Score 238.2; DB 10; Length 408;
Best Local Similarity 80.4%; Pred. No. 3.8e-21;
Matches 279; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 493 CTTTCATCTGTCAGCGCCGAGTACCCGGTGTGACGAGGGTGTGACGAGCTGCAGAG 552
DB 408 CTTTCATTATGCTGAGTGGGAGTACCCAGTGTGACGAGGGGACTCGCGACCTGCAGAA 349
QY 553 AGGACTGCGCAAGGGGTGTCGGCTGAGCGCTGCTACGCGGGGTGTGCGGGGGCGC 612
DB 348 AGCCCTGCGCAGGGGACTCATTCATTGAGTCTCTATGACAGATTATCAGGGGTGC 289
QY 613 CGTGGCTTCTCTGCGCAGCACTCTGTGCGCCAGCCCTGCGGCGCGCTGCGCGTCCAG 672
DB 288 GGTAGCTTCTCTGACAGTTCATTGTGCGCTGCGCCCTGGGTGCGCGCTTCCAC 229
QY 673 CTGCTGTGAGTCCCGTGTGCTTAACAGAGAGGGCCCGGCTGTTCGCGGCCCGCGCCGT 732
DB 228 CTGCTTGTGAGTGGGTGTGCTGACGAGAGGGGCCACCGGTTCGCGGCCCGCGCCGT 169
QY 733 GAGCTTCCCTTACCGCGCGCTGCGCTTTCGTGCGCAATCGCGAGAGAGAGCGCGCT 792
DB 168 GACCTTCCCGCGCGGATTCGGTCCCTTGTGCGGAGCGCGAGAGCGCGCGCT 109
QY 793 GCTGTACAGAGGACAACTGCGCCAGGTGCGCTGAGGGTGCCTCC 839
DB 108 ACTCTCAAGAAGACACAACTGGCTCAGGTGGCTGAGGCCAGGCC 62

RESULT 14

EX636992/c
LOCUS
DEFINITION
BX636992 pBluescript Lion Mus musculus cDNA clone LiONp462F07402
3', mRNA sequence.

assistance we gratefully acknowledge.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES

source
Location/Qualifiers
1..445
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F930003104"
/tissue_type="inner ear"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult inner ear"

ORIGIN

Query Match 23.9%; Score 231.6; DB 13; Length 445;
Best Local Similarity 79.7%; Pred. No. 2.5e-20;
Matches 298; Conservative 0; Mismatches 74; Indels 2; Gaps 2;

Qy 459 CACAGACAGACATCGGCCATCGGTGTGACAGCCTTCATCATGCTGAGCGCGGAGTACC 518
Db |||||
1 CACAGACAGACATTTGGGCTATTGGTGTAAACAGCCTTCATTATGCTGAGTGGGAGTACC 60
Qy 519 CGGTGACAGCAGCGGGTGCACGCCACCTGCAGAGAGGACTCGGCAAGGGGCTGTCCGGC 578
Db |||||
61 CAGAGAGCAGCGAGGG-ACTCGTGACCTTCAGAAAGGCTTGGGCAAGGACTCATTCGTT 119
Qy 579 TGAGCCGCTGTACCGGGGCTGTCCGGGGCGCCGTTGCGCTTCCTCGCGAGCAGCTCTGT 638
Db |||||
120 TGAGTCGCTGTATGACAGATTATCAGGAGGAGCGGTAGCCTTCCTGCAGATTTCATTAT 179
Qy 639 GCGCCAGCCCTGAGCGCGCCCTGCGCTCCAGCTGCCTGCGCCGCTGGCTAACAG 698
Db |||||
180 TTGCTCAACCTGAGGCGCCGCTGCGCTTCACCTGCTTGCATGCGGGTGGCTGACAG 239
Qy 699 AGGAGGCGCCGGCTGTTCGGGCGCCCGCCGTGACCTTCCTCCCTAACG-CGCGGCTGCGC 757
Db |||||
240 AGGAGGCTCCACCGGCTCCCGGCCACGCCGCTAACCTTCCCCACCGTGGCGGCTGCGC 299
Qy 758 GTCTTCGTGCGCAATCGCGAGAGAGCGCGCTGTGTACAGAGGACACACCTGGCC 817
Db |||||
300 GCCTTGTGCGGAGCGCGAGAGCGGCTGCGCTTCCTCTACAAGAAGCATACCTGGCC 359

Qy 818 CAGGTGCGCTGAGG 831
Db |||||
360 CAGGTGCGCTGAGG 373

Search completed: September 19, 2004, 22:07:15
Job time : 2840.1 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:09:24 ; Search time 34.5752 Seconds
(without alignments)
6179.453 Million cell updates/sec

Title: US-10-077-130-4 COPY 16862 17246

Perfect score:

Sequence: 1 ccgaggcgagatcttgaca.....acctgcagcacctggagcgc 385

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs. 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

| | |
|----------------|-------------------|
| Minimum DB seq | length: 9 |
| Maximum DB seq | length: 200000000 |

post-processing: Minimum Match 0%

Post-processing: Minimum Match 0%
Maximum Match 100%

Maximum number 1000
Listing first 45 summaries

Database : Issued Patents NA: *

Database : issued_facilities NA:
1: /cqn2 6/ptodata/2/ina/5A COMB.seg:*

```
1: /cgnz_6/produca/z/tna/cn_COMB.seq:
2: /cgnz_6/produca/2/tna/5R_COMB.seq: *
```

```

2: /cgn2_6/prodata/z/ina/3B_COMB.seq:
3: /cgn2_6/prodata/z/ina/6A_COMB.seq:

```

```
3: /cgn2_6/prodata/z/ina/6A_COMB.seq:
4: /cgn2_6/prodata/z/ina/6B_COMB.seq:*
```

```
4: /eghz_6/ptodata/2/ina/6B_COMB.seq:
5: /cqn2_6/ptodata/2/ina/PCNTUS_COMB.seq:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description | |
|------------|-------|-------------|---------|----|----------------------|-------------------|--|
| | | | | | | | |
| C 1 | 38.8 | 10.1 | 723 | 4 | US-09-252-991A-5281 | Sequence 5281, Ap | |
| C 2 | 38.8 | 10.1 | 744 | 4 | US-09-252-991A-5345 | Sequence 5345, Ap | |
| C 3 | 38.8 | 10.1 | 954 | 4 | US-09-252-991A-5307 | Sequence 5307, Ap | |
| | 37.2 | 9.7 | 1200 | 4 | US-09-507-794A-22 | Sequence 22, Appl | |
| 5 | 37.2 | 9.7 | 1200 | 4 | US-09-505-125A-22 | Sequence 22, Appl | |
| 6 | 37.2 | 9.7 | 1200 | 4 | US-09-502-775A-22 | Sequence 22, Appl | |
| 7 | 37 | 9.6 | 4080 | 4 | US-09-016-434-1342 | Sequence 1342, Ap | |
| C 8 | 35 | 9.1 | 523 | 4 | US-09-621-976-1358 | Sequence 1358, Ap | |
| C 9 | 34.8 | 9.0 | 486 | 4 | US-09-252-991A-1276 | Sequence 1276, Ap | |
| 10 | 34.6 | 9.0 | 2028 | 4 | US-09-252-991A-13079 | Sequence 13079, A | |
| C 11 | 34.6 | 9.0 | 3525 | 4 | US-09-252-991A-12708 | Sequence 12708, A | |
| 12 | 34.6 | 9.0 | 3906 | 4 | US-09-252-991A-13251 | Sequence 13251, A | |
| 13 | 34.4 | 8.9 | 855 | 4 | US-09-252-991A-9597 | Sequence 9597, Ap | |
| C 14 | 34.4 | 8.9 | 1080 | 4 | US-09-252-991A-9553 | Sequence 9553, Ap | |
| 15 | 34.4 | 8.9 | 1098 | 4 | US-09-252-991A-13156 | Sequence 13156, A | |
| C 16 | 34.4 | 8.9 | 1101 | 4 | US-09-252-991A-12790 | Sequence 12790, A | |
| C 17 | 34.4 | 8.9 | 1131 | 4 | US-09-252-991A-9599 | Sequence 9599, Ap | |
| 18 | 34.2 | 8.9 | 4403765 | 3 | US-09-103-840A-2 | Sequence 2, Appl | |
| C 19 | 34.2 | 8.9 | 4411529 | 3 | US-09-103-840A-1 | Sequence 1, Appl | |
| C 20 | 34 | 8.8 | 430 | 4 | US-09-621-976-16656 | Sequence 16656, A | |
| 21 | 34 | 8.8 | 1909 | 3 | US-09-100-193-6 | Sequence 6, Appl | |
| C 22 | 34 | 8.8 | 4112 | 1 | US-08-340-203A-2 | Sequence 2, Appl | |
| C 23 | 34 | 8.8 | 4112 | 2 | US-08-452-567-2 | Sequence 2, Appl | |
| C 24 | 34 | 8.8 | 4112 | 2 | US-08-452-427-2 | Sequence 2, Appl | |
| C 25 | 34 | 8.8 | 4112 | 3 | US-09-085-407-2 | Sequence 2, Appl | |
| 26 | 34 | 8.8 | 4287 | 1 | US-08-244-189-1 | Sequence 1, Appl | |
| C 27 | 34 | 8.8 | 4287 | 1 | US-08-306-691B-53 | Sequence 53, Appl | |

RESULT 1

```

US-09-252-991A-5281/C
; Sequence 5281; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5281
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5281

```

ALIGNMENTS

| | | | | | | |
|------|------|-----|---------|---|---------------------|---------------------|
| C 28 | 34 | 8.8 | 4616 | 1 | US-08-340-203A-1 | Sequence 1, Appli |
| C 29 | 34 | 8.8 | 4616 | 2 | US-08-452-567-1 | Sequence 1, Appli |
| C 30 | 34 | 8.8 | 4616 | 3 | US-08-452-421-1 | Sequence 1, Appli |
| C 31 | 34 | 8.8 | 4616 | 3 | US-09-085-407-1 | Sequence 1, Appli |
| C 32 | 33.8 | 8.8 | 3150 | 3 | US-08-943-768-1 | Sequence 1, Appli |
| C 33 | 33.8 | 8.8 | 3150 | 4 | US-09-865-960-1 | Sequence 1, Appli |
| C 34 | 33.6 | 8.7 | 1173 | 4 | US-09-398-858-13 | Sequence 13, Appli |
| C 35 | 33.6 | 8.7 | 1425 | 4 | US-09-266-965-63 | Sequence 63, Appli |
| C 36 | 33.6 | 8.7 | 53500 | 4 | US-09-266-965-76 | Sequence 76, Appli |
| C 37 | 33.2 | 8.6 | 2407 | 4 | US-09-023-655-812 | Sequence 812, App |
| C 38 | 33.2 | 8.6 | 2793 | 1 | US-08-309-747-1 | Sequence 1, Appli |
| C 39 | 33.2 | 8.6 | 2793 | 1 | US-08-458-298-1 | Sequence 1, Appli |
| C 40 | 33 | 8.6 | 1935 | 3 | US-09-495-050A-150 | Sequence 150, Appli |
| C 41 | 33 | 8.6 | 4403765 | 3 | US-09-103-840A-2 | Sequence 2, Appli |
| C 42 | 33 | 8.6 | 4411529 | 3 | US-09-103-840A-1 | Sequence 1, Appli |
| C 43 | 32.8 | 8.5 | 702 | 4 | US-09-489-039A-1899 | Sequence 1899, App |
| C 44 | 32.6 | 8.5 | 83450 | 4 | US-09-811-469-3 | Sequence 3, Appli |
| C 45 | 32.4 | 8.4 | 474 | 2 | US-08-403-852D-14 | Sequence 14, Appli |

Query Match 10.1%; Score 38.8; DB 4; Length 723;

| | Best Local Similarity | 47.2%; | Pred. No. 0.51; | Mismatches | 132; | Indels | 0; | Gaps | 0; |
|-----|--|--------|-----------------|------------|------|--------|----|------|----|
| | Matches | 118; | Conservative | 0; | | | | | |
| 25 | CCTGCTCACCGGTGACTACCTGCCCCCTAGGGGCTGAGCAGGATGCCATCAGCTGCGGGA | 84 | | | | | | | |
| 542 | CGAGATCGTCGGGACAGCGCCTCCGGGGCGATGGCCACGCGCCGCCACCCGCG | 483 | | | | | | | |
| 85 | AGGCCAGATATGGAGGTCTCGATGACGCCACCCACTGGCTGGCTGTTCGCGACCAA | 144 | | | | | | | |
| 482 | CTTCAGCAGGTAGTAGAGATGCTCGAACCCCTGGGCCATAGCGCAGCGCCTCCGCCCA | 423 | | | | | | | |
| 145 | GCACCAAGTCCAGCCCTCAGCGCAGGGCTGGGTGTCACCACTACCTGGACAGAG | 204 | | | | | | | |
| 422 | GCCCTTGGCTCGCGCCCCAGCGCGCCGACGGCTCGGGTCCGAGCGGTATCGAGACGC | 363 | | | | | | | |
| 205 | GCTCAAGTGTTCACCTGATGGGGGCGCGCTGAGGCCCTGAGTTCCTGGGGAGGCTGT | 264 | | | | | | | |
| 362 | GCTCTCTGCAACAGCCGCGCGCGCGCTGGGCCAGGCCGGTTTCTTGCGCCAGGCGCG | 303 | | | | | | | |
| 265 | GTCTGAAGAC | 274 | | | | | | | |
| 302 | GGCTGACCAC | 293 | | | | | | | |

RESULT 2

US-09-252-991A-5345/c
; Sequence 5345, Application US/09252991A

Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 5345
LENGTH: 744
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5345

Query Match 10.1%; Score 38.8; DB 4; Length 744;
Best Local Similarity 47.2%; Pred. No. 0.52; Mismatches 132; Indels 0; Gaps 0;
Matches 118; Conservative 0;
QY 25 CGTGGTACCGCTGACTACCTGCGCCCTAGGCGGTGAGCAGGATGCCATCAGCTGCGGGA 84
Db 573 CGAGATCGTCGGGACCAAGCGCTCCGGGGCGATGGCCAGCGGAGCGCCACCGCG 514
QY 85 AGGCAGATGTGGAGTCTGATGTCAGCCAGCCAGCCAGCTGCGTGGCTTGTCCGACCAA 144
Db 513 CTTCCAGCAGGTAGTAGAGTCTCGAACCCCTGGCCATAGCGAGCGCTCCGCGCCCA 454
QY 145 GCCCACCACAGTCCAGCCCTCAGCGGCGGTGGTGTACCCAGCTACCTGACAGGAG 204
Db 453 GCCCTGCTCGCGGCCAGCGCCGCGAGCGCTGGGTGCGAGCGGTATGACGACGCG 394
QY 205 GCTCAAGTGTGCTGAGTGGGGCGGTGAGGCGCCCTGAGTTCCTGGGAGGCTGT 264
Db 393 GCTCTGCAACAGCGCGCGCGCGCTGGGCGAGCGCGGTTCCTGGCGCAGGCGCG 334
QY 265 GTCTGAAGAC 274
Db 333 GGCTGACCAC 324

RESULT 3
US-09-252-991A-5307/c
Sequence 5307, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 5307
LENGTH: 954
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5307

Query Match 10.1%; Score 38.8; DB 4; Length 954;
Best Local Similarity 47.2%; Pred. No. 0.55; Mismatches 118; Conservative 0;
Matches 118; Conservative 0;
QY 25 CGTGGTACCGCTGACTACCTGCGCCCTAGGCGGTGAGCAGGATGCCATCAGCTGCGGGA 84
Db 820 CGAGATCGTCGGGACCAAGCGCTCCGGGGCGATGGCCAGCGGAGCGCGCCACCGCG 761

QY 85 AGGCAGTATGTGGAGTCTCGATGTCAGCCAGCCAGCCAGCTGCGTGGCTTGTCCGACCAA 144
Db 760 CTTCCAGCAGGTAGTAGAGTCTCGAACCCCTGGCCATAGCGAGCGCTCCGCGGCA 701
QY 145 GCCCACCACAGTCCAGCCCTCAGCGGCGGTGGTGTACCCAGCTACCTGGAACAGGAG 204
Db 700 GCCCTGCTCGCGGCCAGCGCCGCGAGCGCTGGGTGCGAGCGGTATGACGACGCG 641
QY 205 GCTCAAGTGTGCTGAGTGGGGCGCGCTGAGCCCTGAGTTCCTGGGAGGCTGT 264
Db 640 GCTCTGCAACAGCGCGCGCGCGCTGGGCGAGCGCGGTTCCTGCGCAGGCGCG 581
QY 265 GTCTGAAGAC 274
Db 580 GGCTGACCAC 571

RESULT 4

US-09-907-794A-22
Sequence 22, Application US/09907794A
Patent No. 6635468
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavini, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,794A
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29

```
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 22
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-907-794A-22

Query Match
Best Local Similarity 9.7%; Score 37.2; DB 4; Length 1200;
Matches 125; Conservative 0; Mismatches 123; Indels 4; Gaps 1;

Qy 122 CTGGCTGGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGGCTGGTG 181
Db 715 CTGGGTGGCGGAGGAGCCAGATCCCGAGGAGGACCTGAGGGCCCGAAGCATC 774

Qy 182 TCACGAGCTACCTGGACAGGAGGCTCAAGCTGTCACTGAGTGGGGCGCGTGAAGCC 241
Db 775 CGAGCCCCAGCTGGGA---AGGGCAGCGCGGTGCCAGGGCGGCTGGCACAGTGC 830

Qy 242 CTTGAGTTCCTGGGAGGCTGTCTGAGAGCAATACAGGCAAGGCTGAGCTCTGTG 301
Db 831 CCCCTTCGCGAGCGGGTGGCAGGCCCTCGAGAGGAACCTGAGTGTACCCCTGATCTCAGGC 890

Qy 302 ATCCAGGAGCTGCTGAGTTCCTGAGCAGGCGCTTCGTGGAGGAGCTGAGTTCCTGAGAGC 361
Db 891 CACCAGCCTCTGCGGGCTCCAGCGGGCTCTCTGAAGCCCGCTGAAGAGTTCAGGACTG 950

Qy 362 CACCACCTGCAG 373
Db 951 AAGGCTTGCAG 962

RESULT 5
US-09-905-125A-22
; Sequence 22, Application US/09905125A
; Patent No. 6654376
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-902-775A-22

Query Match          9.7%; Score 37.2; DB 4; Length 1200;
Best Local Similarity 49.6%; Pred. No. 1.5;
Matches 125; Conservative 0; Mismatches 123; Indels 4; Gaps 1;

QY 122 CTGGCTGGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGCGCTGGGTG 181
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 715 CTGGGTGGCGGAGGGAGCCAGATCCCCGAGGAGGACCTTGAGGGCCGCGAAGCATC 774
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 182 TCACCAAGCCTACCTGGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGCGCTGAGGCC 241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 775 CGAGCCCCCAGCTGGGA----AGGGCAGGCGCGTGCCTCCAGGGGCGGTGGCACAGTGC 830
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 242 CCTGAGTTCCTGGGAGGCTGTGTCTGAAGCGAATACAAGCAGAGGCTGAGCTCTGTG 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 831 CCCCTCCGACGGGTGGCAGGCCCTTGAGAGAACTGAGTGTCACTGATCTCAGGC 890
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 302 ATCCAGAGCTGTGAGTCTTGAGCAGGCTTCGTGGAGGAGCTGCAGTTCCTGCAGAGC 361
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 891 CACCAGCTCTGCGGCTCCAGCGGGCTCTCTGAAGCCGCTCTGAAAGCTCAGCGACTG 950
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 362 CACCACCTGCAG 373
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 951 AGGCGCTTGCAG 962
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
US-09-016-434-1342
; Sequence 1342, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1342:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4080 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:

```

LIBRARY: GENBANK
CLONE: G34764
US-09-016-434-1342

Query Match 9.6%; Score 37; DB 4; Length 4080;
Best Local Similarity 46.4%; Pred. No. 2.2;
Matches 121; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 95 GTGAGAGTCTCGATGAGCCACCCACTGGCGTGTTCGCCACCAAGCCACCAAG 154
DB 263 GTGAGAGCCCGCGCCCTACACCTCCGACACCGGATCCGGCCACCGCGCCAG 322
QY 155 TCCAGCCCTCAGCGAGGGGTGGGTGTACACCCCTACCTGTGACAGAGAGGTCTCAAGCTG 214
DB 323 CCGTAAAGGGTCTGAAGCGCGGCGCACCGCTCGCCGACAGGTCTATGAGGGCGGCTC 382
QY 215 TCACCTGAGTGGGGCGCGCTGAGGCGCCCTGAGTTCCTGGGAGGCTGTCTGAGAC 274
DB 383 GCACCCAACTGAGCGCGGAGGAGCAACGCCAGCGCCGCGCGCGGCGGAGGGC 442
QY 275 GAATACAGGCAAGCTGAGCTCTGTGATCCAGGAGTGTCTGAGTTCTGAGCAGGCTTC 334
DB 443 AACCGCACCGCGACCCCGCGCGCAACGAGGCCCTTGGCGCGGTGAGGTGGCGGTG 502
QY 335 GTGAGGAGCTGCAGTTCCTG 355
DB 503 CTGTGCTCATCTGCTCCTG 523

RESULT 8

US-09-621-976-1358/c
; Sequence 1358, Application US/09621976
; Patent No. 8639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1358
; LENGTH: 523
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 176..421
; NAME/KEY: sig_peptide
; LOCATION: 176..268
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 10.8999996185303
; OTHER INFORMATION: seq LLLALSLAHLCA/FS
US-09-621-976-1358

Query Match 9.1%; Score 35; DB 4; Length 523;
Best Local Similarity 46.1%; Pred. No. 4.4;
Matches 113; Conservative 1; Mismatches 131; Indels 0; Gaps 0;

QY 19 CATCTAGCTGTACCGCTGACTACCTGCCCCCTAGGGGCTGAGCAGGATGCCATCAGCT 78
DB 290 CACCTGAGGTTCTAGGCTAAAGCCCGACGATAGTGGCAGCTGAGAGGGCCAGGAGGA 231
QY 79 GCGGAGGCGCAGTATGTGGAGTCTTGGATGAGCCACCCACTGGCTGGCTGTTCGG 138
DB 230 GCTGAGGCTCAGGGGTGGGGATAGCGAAGGAAGTAGAAGTGGTCTGGCATGTGCC 171
QY 139 CACCAAGCCCAAGTCCAGCCCTCAGCGCAGGGTGGTGTCTACAGCCCTACCTGGA 198
DB 170 CAGTTCCACCCACCTTCCCTCTAGGGAGAGGAGTGGCAGAGCAACACTGAGGC 111

QY 199 CAGGAGGCTCAAGCTGTACCTGAGTGGGGGGCGCTGAGGCCCTGAGTTCCTCTGGGA 258
DB 110 TGCAGGACACAGACACTCCCTATGGGGGTATGCGAGCTCTCGTTCTGTSACT 51
QY 259 GCGTG 263
DB 50 GCGTG 46

RESULT 9

US-09-252-991A-1276/c
; Sequence 1276, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1276
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1276

Query Match 9.0%; Score 34.8; DB 4; Length 486;
Best Local Similarity 50.6%; Pred. No. 4.8;
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 216 CACCTGAGTGGGGCGCGCTGAGGCCCTGAGTTCCTCTGGGAGGCTGTCTGAAGCG 275
DB 266 CGCTCGGGCGTAGGGCAAGGCTTCGGGTGAGGGAACCGCGCTTCGCGAGG 207
QY 276 AATACAGGCAAGGCTGAGCTCTGTGATCCAGGAGCTGTGAGTTCGAGCAGGCTTCG 335
DB 206 ATTTCGCGCGCGCGCCATTTCGCTCCAGTTCTGTCATCAGTGGCTGTGCCCCG 147
QY 336 TGGAGGAGTGCAGTTCCTGACAGCCACCACCTCAGCACCTGGA 381
DB 146 AGAAGCTGTTGAGGACCGGAGCTCGGCGCTTCAGGGCCTGGA 101

RESULT 10

US-09-252-991A-13079
; Sequence 13079, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13079
; LENGTH: 2028
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13079

Query Match 9.0%; Score 34.6; DB 4; Length 2028;
Best Local Similarity 52.4%; Pred. No. 7.5;
Matches 76; Conservative 0; Mismatches 69; Indels 0; Gaps 0;


```
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9553
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9553

Query Match      8.9%; Score 34.4; DB 4; Length 1080;
Best Local Similarity 48.0%; Pred. No. 7.3; Indels 0; Gaps 0;
Matches 98; Conservative 0; Mismatches 106;

QY 11 ATCTTTGACATCTAGTGTGTCACCGCTGACTACCTGCCCCCTAGGGGCTGAGCAGGATGCC 70
   |||||
Db 89 ACCTTCGTGCGCATCGCGGACACGGTGTTCACCCGCGCGCGGAGGGGTGAATCGC 148

QY 71 ATCAGCTGCGGGAAGGCAGTATGTGGAGTCTCTGATGAGCCACCCACTCGCGCTGG 130
   |||||
Db 149 ACCCAGTCGCGCGTCAAGTATGATGAAGCGGCTCGAAGAGGACGTCTCTGCAGCGCTCG 208

QY 131 CTTGTCCGCACCAAGCCACCAAGTCCAGCCCTCAGCGGAGGCTGGGTGTCAACAGCC 190
   |||||
Db 209 CTGTTGAGCGCGGACGAGGACACCGCTCACCGCCGAGGCCAGGTGCTCTCGGC 268

QY 191 TACTTGACAGAGGCTCAAGCTG 214
   |||||
Db 269 TATGCCGCGGATCTCAAGCTG 292

RESULT 15
US-09-252-991A-13156/c
; Sequence 13156, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107156.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13156
; LENGTH: 1098
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13156

Query Match      8.9%; Score 34.4; DB 4; Length 1098;
Best Local Similarity 57.4%; Pred. No. 7.3;
Matches 62; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 273 ACGAATCAAGCAAGGCTGAGTCTGTGATCCAGGAGCTGCTGAGTCTGAGCAGGCT 332
   |||||
Db 911 ACGAAGCGCTTGTCCGGCCGATACTGCGATCCAGGCGTTGAGGCTCCGTCAGCTCC 852

QY 333 TCGTGGAGAGTGCAGTCTCTGAGAGCCACCCACCTGACCACTGG 380
   |||||
Db 851 TCGAGCAGGTCGGCAGTCTCTCCAGGCAAGCGCCAGGTCGGCAGG 804

Search completed: September 19, 2004, 22:12:45
Job time : 43.5752 secs
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:07:04 ; Search time 1122.53 Seconds
(without alignments)
10242.013 Million cell updates/sec

Title: US-10-077-130-4_COPY_16862_17246

Perfect score: 385

Sequence: 1 ccgaggagagatcttgaca.....acctgcagcactggagcgc 385

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rtd:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 110.6 | 28.7 | 512 | 12 | BJ059798 |
| 2 | 83 | 21.6 | 734 | 14 | CF532602 |
| C 3 | 74.8 | 19.4 | 528 | 12 | BG792090 |
| C 4 | 74.8 | 19.4 | 532 | 12 | BG791767 |

| | | | | | |
|----|------|------|------|----|----------|
| 5 | 43 | 11.2 | 507 | 9 | AV388253 |
| 6 | 43 | 11.2 | 925 | 29 | CNS0091P |
| 7 | 42.6 | 11.1 | 965 | 13 | BUS00169 |
| 8 | 42 | 10.9 | 428 | 9 | AUI08162 |
| 9 | 40.8 | 10.6 | 881 | 13 | BQ923183 |
| 10 | 40.4 | 10.5 | 628 | 12 | BI527461 |
| 11 | 40 | 10.4 | 919 | 12 | BI159893 |
| 12 | 39.8 | 10.3 | 846 | 29 | CNS010RQ |
| 13 | 39.6 | 10.3 | 935 | 29 | CNS006XK |
| 14 | 39.4 | 10.2 | 303 | 10 | BF771180 |
| 15 | 39.4 | 10.2 | 347 | 28 | AQ469228 |
| 16 | 39.4 | 10.2 | 354 | 10 | BF893776 |
| 17 | 39.4 | 10.2 | 390 | 10 | BF772502 |
| 18 | 39.4 | 10.2 | 390 | 12 | BI060781 |
| 19 | 39.4 | 10.2 | 393 | 10 | BF926934 |
| 20 | 39.4 | 10.2 | 444 | 10 | AW484180 |
| 21 | 39.4 | 10.2 | 612 | 12 | BG337262 |
| 22 | 39.4 | 10.2 | 710 | 12 | BI823073 |
| 23 | 39.4 | 10.2 | 715 | 14 | CA778105 |
| 24 | 39.4 | 10.2 | 825 | 13 | BUS28667 |
| 25 | 39.4 | 10.2 | 849 | 12 | BM018737 |
| 26 | 39.4 | 10.2 | 851 | 12 | BG469835 |
| 27 | 39.4 | 10.2 | 873 | 13 | BX357518 |
| 28 | 39.4 | 10.2 | 881 | 13 | BQ980589 |
| 29 | 39.4 | 10.2 | 884 | 12 | BI757129 |
| 30 | 39.4 | 10.2 | 903 | 13 | BQ923314 |
| 31 | 39.4 | 10.2 | 910 | 12 | BI757166 |
| 32 | 39.4 | 10.2 | 925 | 13 | BUS28516 |
| 33 | 39.4 | 10.2 | 933 | 13 | BQ986480 |
| 34 | 39.4 | 10.2 | 955 | 13 | BUI75005 |
| 35 | 39.4 | 10.2 | 994 | 13 | BX353043 |
| 36 | 39.4 | 10.2 | 995 | 13 | BQ673197 |
| 37 | 39.4 | 10.2 | 1015 | 13 | BQ058429 |
| 38 | 39.4 | 10.2 | 1201 | 13 | BX417636 |
| 39 | 39.4 | 10.2 | 1214 | 13 | BUI57066 |
| 40 | 39.2 | 10.2 | 1163 | 13 | BX377844 |
| 41 | 39.2 | 10.2 | 1201 | 13 | BX377844 |
| 42 | 39 | 10.1 | 816 | 14 | CK139732 |
| 43 | 39 | 10.1 | 1201 | 29 | CNS016BR |
| 44 | 38.8 | 10.1 | 273 | 10 | AW789770 |
| 45 | 38.8 | 10.1 | 401 | 14 | HI9853 |

ALIGNMENTS

RESULT 1
BJ059798
LOCUS
DEFINITION
BJ059798 NIBB Mochii normalized Xenopus tailbud library Xenopus laevis cDNA clone XL063014 5', mRNA sequence.
ACCESSION
BJ059798
VERSION
BJ059798.1
KEYWORDS
EST.
SOURCE
Xenopus laevis (African clawed frog)
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 512)
Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.

```

http://xenopus.nibb.ac.jp.
Location/Qualifiers
FEATURES
  source
    1..512
      /organism="Xenopus laevis"
      /mol_type="mRNA"
      /db_xref="taxon:8355"
      /clone="XL063014"
      /tissue_type="whole embryo"
      /dev_stage="stage 25"
      /clone_lib="NIBB Mochii normalized Xenopus tailbud
      library"

ORIGIN
Query Match      28.7%; Score 110.6; DB 12; Length 512;
Best Local Similarity 71.2%; Pred. No. 3.2e-14;
Matches 146; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 179 GTGTCCAGAGCTACCTGAGAGAGGCTCAAGCTGTCACTGAGTGGGGGGCGCTGAG 238
Db 28 GGGTCTACCGCTTACTTGGATAAGAGACTGAAGCTATCATCTGAGTGGGCTGCAGGAGAA 87
QY 239 GCCCTGAGTTCCCTGGGGAGGCTGTGTCGAAGACGAATACAAGGCAAGGCTGAGCTCT 298
Db 88 GCACAGAACTTCTGGAGTGTGTATCAGAGAGGAGTACAGAGAGAGCTCAGATT 147
QY 299 GTGATCCAGGAGCTGCTGAGTTTCAGAGAGCGCTTCGTGGAGGAGCTGCAGTTCTCTCAG 358
Db 148 CTGATCCAGGAGCTGCTAAACACAGAGAGGAGTACGTGAAGGAGCTACAGTTCTTTCAG 207
QY 359 AGCCACCACTCCGACCTCGAGC 383
Db 208 AACCATCACTACATCACTGAGC 232

RESULT 2
LOCUS
DEFINITION
  CF532602 734 bp mRNA linear EST 12-SEP-2003
  UI-M-GH0-cgw-a-11-0-UI.r1 NIH_BMAP_GH0 Mus musculus cDNA clone
  IMAGE:30357226 5', mRNA sequence.
ACCESSION
  CF532602
VERSION
  CF532602.1 GI:34584570
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 734)
  NIH-MGC http://mgi.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
FEATURES
  source
    1..734
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6"
      /db_xref="taxon:10090"
      /clone="IMAGE:30357226"
      /tissue_type="Whole brain"
      /dev_stage="1, 5, and 15 days newborn"
      /lab_host="DH10B (T1 phage resistant)"
      /clone_lib="NIH_BMAP_GH0"
      /note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;

Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CGAAGTGAAT. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."
ORIGIN
Query Match      21.6%; Score 83; DB 14; Length 734;
Best Local Similarity 86.0%; Pred. No. 4.9e-08;
Matches 92; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 35 GCTGACTACTCTCCCTAGGGCTGAGCAGGATGCATCAGCTCGGGAGAGCCAGTAT 94
Db 628 GCTGACTATCTGCCACTGGAGCTGAGCAGGATGCCATCTTCGAGAGAGCCAGTAT 687
QY 95 GTGAGAGTCTCTGGATGCGAGCCACCCACTGCGCTGGCTTGTCCGAC 141
Db 688 GTGAGAGTCTCTGGATGCGAGCCACCCACTGCGCTGGCTTGTCCGAC 734

RESULT 3
LOCUS
DEFINITION
  BG792090 528 bp mRNA linear EST 30-MAY-2001
  UTSW_H24F12 UTSW Adult Mouse Cardiac Muscle Library Mus musculus
  cDNA clone UTSW_H24F12, mRNA sequence.
ACCESSION
  BG792090
VERSION
  BG792090.1 GI:14127660
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 528)
  Gallardo, T.D., Schageman, J.J., Pertsemilidis, A., Garner, H.R.,
  Williams, R.S., and Shchet, R.V.
  UT Southwestern Medical Center, Adult Mouse Cardiac cDNA Library
  Unpublished (2001)
  Contact: Schageman JJ
  Shohet/Garner Labs
  University of Texas Southwestern Medical Center
  6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA
  Tel: 214 648 1674
  Email: Jeff.Schageman@UTSouthwestern.edu
cDNA library constructed by UTSW as a component of the Program for
Genomic Applications (PGA) and the Reynolds Heart Disease
Prevention Grants for use in cDNA microarray experiments. Sequence
Quality: Sequence ends were trimmed based on percentage of ambigu-
us base calls or 'N's in windowed segments. Sequencing: First-pass
sequencing; ABI Prism 377 sequencer and analysis software.
Seq primer: M13/pUC Reverse.
Location/Qualifiers
FEATURES
  source
    1..528
      /organism="Mus musculus"
      /mol_type="mRNA"
      /db_xref="taxon:10090"
      /clone="UTSW_H24F12"
      /sex="Pooled"
      /tissue_type="Cardiac muscle"
      /dev_stage="2 months"
      /lab_host="DH5a"
      /clone_lib="UTSW Adult Mouse Cardiac Muscle Library"
      /note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
      Primer: Oligo dT. RNA isolation: cytoplasmic RNA preps
      (Mannatis); cloning Technique: CUA Cloning (Clontech),

```

Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACGTCCACTGAATTCGAGTG-->. Other
information regarding entire library may be found at
http://pga.swned.edu/Data/Libraries/microarray_cdna_librar
ies.htm."

ORIGIN

Query Match 19.4%; Score 74.8; DB 12; Length 528;
Best Local Similarity 77.4%; Pred. No. 2.8e-06;
Matches 113; Conservative 0; Mismatches 31; Indels 2; Gaps 2;
QY 240 CCCCTGAGTTCCTGGGAGGCTGTCTCTGAAG-ACGAATACAAGGCAAGGCTGAGCTCT 298
|||||
Db 477 CCCCGAGTTCCTGGTGAAGGCTGTCTGAGGNATGAGTATAGAACGAGGCTGAGCTCT 418
QY 299 GTGATCCAGGAGCTGCTGAGTCTGAGCAGGCTTCGTGGAGGAGCTGCGAGTTCCTCGAG 358
|||||
Db 417 GTCATCCAGGAGTGTCTGAGTTCAGAGCAGGCTTTTGTGGTGANCNGCAGTCTCTTGAG 358
QY 359 AGCCA-CCACCTGCAGCACCTGGAGC 383
|||||
Db 357 ANCCACCCACATGAAGCAGCAGTGGAGC 332

RESULT 4

BG791767/c
LOCUS
DEFINITION
UTSW_H15F12 UTSW Adult Mouse Cardiac Muscle Library Mus musculus
cDNA clone UTSW_H15F12, mRNA sequence.

ACCESSION
BG791767
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
Gallardo, T.D., Schageman, J.J., Pertsemilidis, A., Garner, H.R.,
Williams, R.S., and Shohet, R.V.
TITLE
UT Southwestern Medical Center, Adult Mouse Cardiac cDNA Library
JOURNAL
COMMENT
Contact: Schageman JJ
Shohet/Garner Labs
University of Texas Southwestern Medical Center
6050 Harry Hines Blvd., N2.226, Dallas, TX 75390, USA
Tel: 214 648 1674

FEATURES
source
1..532
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="UTSW_H15F12"
/sex="Pooled"
/tissue_type="Cardiac muscle"
/dev_stage="2 months"
/lab_host="DH5a"
/clone_lib="UTSW Adult Mouse Cardiac Muscle Library"
/note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dt. RNA isolation: cytoplasmic RNA preps
(Mannatis); Cloning technique: CUA Cloning (Clontech),
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACGTCCACTGAATTCGAGTG-->. Other
information regarding entire library may be found at
http://pga.swned.edu/Data/Libraries/microarray_cdna_librar
ies.htm."

ORIGIN

Query Match 19.4%; Score 74.8; DB 12; Length 532;
Best Local Similarity 77.4%; Pred. No. 2.8e-06;
Matches 113; Conservative 0; Mismatches 31; Indels 2; Gaps 2;
QY 240 CCCCTGAGTTCCTGGGAGGCTGTCTCTGAAG-ACGAATACAAGGCAAGGCTGAGCTCT 298
|||||
Db 477 CCCCGAGTTCCTGGTGAAGGCTGTCTGAGGNATGAGTATAGAACGAGGCTGAGCTCT 418
QY 299 GTGATCCAGGAGCTGCTGAGTCTGAGCAGGCTTCGTGGAGGAGCTGCGAGTTCCTCGAG 358
|||||
Db 417 GTCATCCAGGAGTGTCTGAGTTCAGAGCAGGCTTTTGTGGTGANCNGCAGTCTCTTGAG 358
QY 359 AGCCA-CCACCTGCAGCACCTGGAGC 383
|||||
Db 357 ANCCACCCACATGAAGCAGCAGTGGAGC 332

RESULT 5

AV388253
LOCUS
DEFINITION
AV388253 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
cDNA clone CM024C03_r, mRNA sequence.

ACCESSION
AV388253
VERSION
KEYWORDS
SOURCE
Chlamydomonas reinhardtii
ORGANISM
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

REFERENCE
AUTHORS
TITLE
A large scale structural analysis of cDNAs in a unicellular green
alga, Chlamydomonas reinhardtii. I. Generation of 3433
non-redundant expressed sequence tags
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1..507
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="CM024C03_r"
/dev_stage="photoautotrophic growth"
/clone_lib="Chlamydomonas reinhardtii C9"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

ORIGIN

Query Match 11.2%; Score 43; DB 9; Length 507;
Best Local Similarity 54.0%; Pred. No. 31;
Matches 88; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 122 CTGCGCTGGCTGTCTCGCACCACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGCGTGGGTG 181
|||||
Db 37 CGGCGGTGGCTTCTCGCCAGGCGGCGGTGCGCCCGCGCGTTCGCCGACGGTGTCTC 96
QY 182 TCACAGCTTACCTGAGCAGGAGGCTCAAGCTGTCACTGAGTGGGGGCGCCCTGAGGCG 241
|||||
Db 97 GTCGACGCTGTGTGTCGCGCTTCCACGCCGAGAACCCCTAGGTGGGCTCCGTTGGCC 156
QY 242 CCTGAGTTCCTGGGAGGCTGTCTGAAGCAAGTATCAAGG 284
|||||
Db 157 CTGACTTCAAGCCCGCGCGCGTGTTCGACGAGGATTCAGG 199

REFERENCE
1 (bases 1 to 428)
AUTHORS
Sasaki, T. and Yamamoto, K.
TITLE
Rice cDNA from callus (2000)
JOURNAL
Unpublished (2000)
COMMENT
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel.: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = "RGP".

FEATURES
source
1..428
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6426133"
/clone="C50433"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site 1: Sali; Site 2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the Sali-NotI site
of pBluescript II SK+ phagemid."

ORIGIN
Query Match 10.9%; Score 42; DB 9; Length 428;
Best Local Similarity 47.1%; Pred. No. 47;
Matches 129; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
QY 111 CAGCCACCCACTCGCTGGTGTTCGCCACCAAGCCCAAGTCCAGCCCTCACGGC 170
Db 53 CACCTCGCAGGTCCGAGCCTCTACCCACACCAAGCAGCAGCAAGCGTCGTGC 112
QY 171 AGGCTGGTGTTCACAGCTTACCTGGACAGAGGCTCAAGTCTCACTGAGTGGGG 230
Db 113 TGGCGAGGTATCAGACACCTGAAGGAGCTGAAGCGGACAGCAGCGGATCGCGCG 172
QY 231 CCGCTGAGGCCCTCGAGTTCCTCGGGAGGCTGTCTGAGAGCAATACAAGCAAGGC 290
Db 173 CAGCGCGCGGGGATTAACACGCAACGACGAGGACGACGATCGGTGGGGC 232
QY 291 TGAGCTGTGATCAGAGAGTGTCTGAGTTCAGAGGCTTCTGTGGAGAGCTGAGT 350
Db 233 GCGGTTCGGCGCGCGCAGAGTGTCTGCGCAGCGAGCGGACGAGCTGCGGTGAGC 292
QY 351 TCCTGCAGAGCCACACCTGCGACACCTGGAGCG 384
Db 293 CGCGGTGAGCGCGGAGGAGTCTGTGTGCG 326

RESULT 9
BQ923183
LOCUS
DEFINITION
AGENCOURT_8803577 NIH_MGC_101 Homo sapiens cDNA clone IMAGE:6426133
5' mRNA sequence.
ACCESSION
BQ923183
VERSION
BQ923183.1 GI:22339214
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 881)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2609 row: a column: 14
High quality sequence stop: 645.

FEATURES
source
1..881
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6426133"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="PH10B (phage-resistant)"
/clone_lib="NIH_MGC_101"
/note="Organ: lung; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Query Match 10.6%; Score 40.8; DB 13; Length 881;
Best Local Similarity 51.7%; Pred. No. 1.2e+02;
Matches 93; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 18 ACATCTAGCTGTCACCGCTGACTCTCCCTAGGCGCTGAGCAGGATGCATCACGC 77
Db 576 AGCCTTCTGTGACGCGCTAAGTGCAGCTCTCTTTGCCCTCAAGGTGCTCAACATGA 635
QY 78 TGGCGGAAGCCAGTATGTGGAGGTCTCGATGCGCCACCCACCTGCGCTGTGTC 137
Db 636 TGCCGAGAGAGCTGTTGAGGCTGCTGCGCCACCCAGAGAAACAGAGAAAGCCC 695
QY 138 GCACCAAGCCCAAGTCCAGCCCTCTACGCGAGGCTGGGTGTCCACGAGCTACCTGG 197
Db 696 TGGAGAAGTGTCTCCCGGCTCTCTTGAGGCTGCTGAGCTGGTGCGACCTCCCTGG 755

RESULT 10
B1527461
LOCUS
DEFINITION
1024081009.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION
B1527461
VERSION
B1527461.1 GI:15368035
KEYWORDS
EST.
SOURCE
Chlamydomonas reinhardtii
ORGANISM
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
REFERENCE
1 (bases 1 to 628)
AUTHORS
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C.,
Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.
TITLE
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1024b
JOURNAL
Unpublished (2001)
COMMENT
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
source
1..628
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"

Job time : 1127.53 secs

```

Db      4  ACGCCTTCGTGAGCAGCGCTAAGTGCCAGCTCCTCTTTGCCCCCAAGGTGCTCAACATGA 63
QY      78  TCGCGGAAGGCGAGTATCTGGAGGTCTCTGGATGCGCCACCCACCTGCGCTGGCTTGTCC 137
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      64  TCCCGGAGAGAGCTGTTGAGGCCCTGGCTGAGCCACCGAGAAGCAGAGAGGCC 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      138  GCACCAAGCCACCAGTCCAGCCCTCACGGCAGGGCTGGGTGTACACGACCTACC 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      124  TGGAGAAGTTGCTCCCGCCCTCTCTTGAGGGCTGCCTGAGCTGGTGGCACCCCTCCC 180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AQ469228/c
LOCUS   AQ469228               347 bp    DNA        linear    GSS 23-APR-1999
DEFINITION  CITBI-E1-2601M16.TR CITBI-E1 Homo sapiens genomic clone 2601M16,
GENOMIC SURVEY SEQUENCE.
ACCESSION  AQ469228
VERSION    AQ469228.1  GI:4653118
KEYWORDS   GSS.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 347)
AUTHORS    Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
TITLE      Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
JOURNAL    Unpublished (1997)
COMMENT    Other GSSs: CITBI-E1-2601M16.TF
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source            1..347
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /clone="2601M16"
                        /sex="male"
                        /cell_type="sperm"
                        /clone_lib="CITBI-E1"
                        /note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
                        Caltech Human BAC Library D"

ORIGIN
Query Match      10.2%; Score 39.4; DB 28; Length 347;
Best Local Similarity 51.4%; Pred.No.1.6e+02;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY      18  ACATCTACGTGGTCACCGCTGACTACCTGCGCCCTAGGGGCTGAGCAGGATGCCATCAGC 77
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      194  ACGCTTCGTGAGACCGGCTAAGTGCCAGCTCTCTTTGCCCTCAAGTGCTCAACATGA 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      78  TCGCGGAAGGCCAGTATGTGGAGTCTCTGGATGCGCCACCCACCTGCGCTGGCTTGTCC 137
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      134  TGCCCGAGGAGAGCTGGTTGAGGCCCTGGCTGCGCACCGCAGAGAAGAGAGGCC 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      138  GCACCAAGCCACCAGTCCAGCCCTCACGGCAGGGCTGGGTGTACACGACCTACC 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      74  TGGAGAAGTTGCTCCCGCCCTCTCTTTGAGGGCTGCCTGAGCTGGTGGCACCCCTCCC 18
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:07:04 ; Search time 3119.75 Seconds
(without alignments)
10242.013 Million cell updates/sec

Title: US-10-077-130-4_COPY1_1070

Perfect score: 1070
Sequence: 1 tgcctaccagagccacac.....tagtgcgcgagccgcggtt 1070

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rod:*
 - 26: em_gss_phg:*
 - 27: em_gss_vrl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| C 1 | 260 | 24.3 | 911 | 13 | BU180714 |
| C 2 | 249.8 | 23.3 | 728 | 29 | AG034556 |
| C 3 | 188 | 17.6 | 945 | 29 | CNS04AD2 |
| C 4 | 169.4 | 15.8 | 2650 | 11 | AK031074 |

| | | | | | |
|------|-------|------|------|----|----------|
| 5 | 168.8 | 15.8 | 904 | 13 | BU144883 |
| 6 | 167 | 15.6 | 5558 | 11 | AK029863 |
| 7 | 159.6 | 14.9 | 2710 | 11 | AK034852 |
| 8 | 147.8 | 13.8 | 496 | 14 | CA368743 |
| 9 | 144.8 | 13.5 | 752 | 29 | CNS01YES |
| C 10 | 127.4 | 11.9 | 539 | 9 | AL927364 |
| 11 | 124.2 | 11.6 | 1136 | 12 | BM799726 |
| 12 | 122.4 | 11.4 | 929 | 13 | BO679064 |
| 13 | 122.2 | 11.4 | 1017 | 13 | BO679833 |
| 14 | 119 | 11.1 | 910 | 12 | BG490878 |
| 15 | 112.8 | 10.5 | 407 | 10 | AM170791 |
| C 16 | 112.2 | 10.5 | 437 | 29 | CE070659 |
| 17 | 111.2 | 10.4 | 718 | 13 | BO770422 |
| 18 | 108.8 | 10.2 | 780 | 29 | CNS04142 |
| 19 | 107.8 | 10.1 | 534 | 14 | CD555386 |
| 20 | 106.8 | 10.0 | 519 | 12 | BI789591 |
| 21 | 106.6 | 10.0 | 518 | 10 | BE648756 |
| 22 | 106.2 | 9.9 | 756 | 14 | CB249772 |
| 23 | 106.2 | 9.9 | 4130 | 11 | AK048144 |
| 24 | 99.6 | 9.3 | 947 | 29 | AG069825 |
| 25 | 98.4 | 9.2 | 848 | 29 | AG032377 |
| C 26 | 98 | 9.2 | 1165 | 29 | AG030649 |
| C 27 | 98 | 9.2 | 1542 | 29 | AG032943 |
| C 28 | 97.6 | 9.1 | 1598 | 29 | AG030579 |
| C 29 | 96.8 | 9.0 | 1625 | 29 | AG043477 |
| 30 | 96.6 | 9.0 | 1516 | 12 | BG809984 |
| 31 | 94.8 | 8.9 | 1026 | 28 | BZ569417 |
| 32 | 94.4 | 8.8 | 932 | 29 | CNS0072Q |
| C 33 | 94.4 | 8.8 | 1610 | 28 | BZ569386 |
| C 34 | 94.2 | 8.8 | 1452 | 29 | AG032979 |
| 35 | 94 | 8.8 | 935 | 29 | CNS006XK |
| 36 | 94 | 8.8 | 982 | 13 | BX415111 |
| C 37 | 92 | 8.6 | 982 | 13 | BX415111 |
| 38 | 92 | 8.6 | 1124 | 29 | AG041123 |
| 39 | 91.8 | 8.6 | 1085 | 29 | AG152796 |
| 40 | 91.2 | 8.5 | 454 | 10 | AW410057 |
| 41 | 91 | 8.5 | 1189 | 29 | AG030608 |
| 42 | 91 | 8.5 | 1341 | 29 | AG030611 |
| C 43 | 90.2 | 8.4 | 863 | 12 | BG809696 |
| 44 | 90.2 | 8.4 | 1070 | 28 | BZ557810 |
| 45 | 90.2 | 8.4 | 1116 | 12 | BG810038 |

ALIGNMENTS

RESULT 1

BU180714/c

LOCUS

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

BU180714

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

BU180714

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

BU180714

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

BU180714

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

BU180714

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

BU180714

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

BU180714

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

BU180714

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

BU180714

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

BU180714

AGENCOURT_7968637 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6011039

5', mRNA sequence.

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE
AUTHORS

Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fzames, C., Wincker, P., Brottier, P., Quetier, P.,
Saurin, W. and Weissenbach, J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence

JOURNAL
MEDLINE
PUBMED

Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645

REFERENCE
AUTHORS

Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fzames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish tetraodon nigroviridis

JOURNAL
MEDLINE
PUBMED

Genome Res. 10 (7), 939-949 (2000)
20359837
10899143

REFERENCE
AUTHORS

Genoscope.
3 (bases 1 to 945)
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT

This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.

FEATURES
source

Location/Qualifiers
1..945
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="095C14"
/clone_lib="G"
/note="Genoscope sequence ID : CORG095B07SP1-end :
PUC-ori"

ORIGIN

Query Match 17.6%; Score 188; DB 29; Length 945;
Best Local Similarity 56.7%; Pred. No. 6.6e-16;
Matches 388; Conservative 1; Mismatches 286; Indels 9; Gaps 2;

QY 78 CAGCCACAGTTACGGGGGCCCGCTTTCTACCCGGGCCCAAGGCTTCGTGTGTCG 137

DB 893 CACACCTATTGGGGAGTCCCMAGGTTTCTAACGGCGCCGGAAGGCTTTTCGTGTGT 834

QY 138 GTGGCAGAGGACGACCCCTCAGCTGCAGATCGTGGGTATCCAGCCACAGTGAGC 197

DB 833 GTGGCAAGATGGGACTCTGAGCTGCACTGTGTGGAAATCCGACCCCGCTGATAACC 774

QY 198 TGGGAGAGGACACGACGCGCGGTGACGCGCGCGCGCTTCCTCTGGGCCAGGACGCG 257

DB 773 TGGGAAGGACAGCTGAAGCTGACGCTGTGGGGAGCATCAAGACCGTGGAGATGGA 714

QY 258 GACCTTACCGCTCCTACTATCTGAGACTGTGGCGTGGCGGACAGTGGCGCAATACGTGTGC 317

DB 713 GATGTGTACCGCTTGCACATCTACGAGCTGACCCCTGGAAGACACGCGTCAGTATATGTGC 654

QY 318 CGCGCGCGCAATGTCATAGGCGAGGCTTCTGCTCCCTGGGCTTCGACAGTGGACGGGAG 377

DB 653 AGAGCCAGACACAGTTGGTGAAGCTGACCGCGCTGACCTCAAGTGGCGCTTCGCG 594

QY 378 CGCGGTGCGCGACGAGGCGCCGCACTTCTGTGTGGCGCCACAGTCCATCCCGTGC 437

DB 593 GCAGAGATGCCCAAGAGGCGCCCTGTCTCTGTCTCAAGCTGCGACCGCGCGTGTAGGC 534

QY 438 GAGGGCTCAGAGGACACCTTCGCTGCGCGGTGGTGGCTCCCGAGCGCGCAGTGAGC 497

DB 533 CTGGAGAGAGCGTGTGCTGTGCTGCGGAGTCCAGCTCACCCCGGACCACTTGAA 474

QY 498 TGGTCCAAAGGACGCGGGCGCGCCCTGGGTGAGCCGACGCGGCCCGCGTGCCTGGAGGAG 557
DB 473 TGGGAGAGGACGCGGACGCTACCTAGGGGAG-----AGCAATCGCGTCAAGATTGTTC 420
QY 558 CTGGCGGAGCAAGTGGCTGGCCATTTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTAC 617
DB 419 GACAGCGACAGCAGCATTCTGAAGATCCAAAGGGGTGGTCACTGTGACGCGCGCACTAC 350
QY 618 GAGGTCCGCGCGCGAGAACCGCTGGGGCGGTGCGAGCGCGCGCGCGCGCGCTAGTGGTGGAC 677
DB 359 ACCTGCGAGGCGCGAGAACACCGCTGGCGCGTGCATGCTGCAGCGCTCTGCAGCGTGGAC 300
QY 678 TCGGAGCGCGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTTGGCGCAC 737
DB 299 GCCAGATTCACAGCCTGATAAAACAGATAAAGCA---CCTCCCTGCTGTCTCAC 243
QY 738 CTGACGCGCGCGCGCGAGGCTATG 761
DB 242 CTTCAAAGCGCAAGAGGAATG 219

RESULT 4
AK031074
LOCUS

DEFINITION

AK031074 2650 bp mRNA linear HTC 18-SBP-2003
Mus musculus 13 days embryo forelimb cDNA, RIKEN full-length
enriched library, clone:5930402G23 product:hypothetical
immunoglobulin and major histocompatibility complex domain
containing protein, full insert sequence.

ACCESSION
VERSION

AK031074.1 GI:26326994
HTC; CAP trapper.

KEYWORDS
SOURCE

Mus musculus (house mouse)
Mus musculus

ORGANISM

Mus musculus

REFERENCE
AUTHORS

High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

JOURNAL
MEDLINE
PUBMED

99279253
10349636

REFERENCE
AUTHORS

Carninci, P. and Hayashizaki, Y.

Prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL
MEDLINE
PUBMED

2049374
11042159

REFERENCE
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, I., Ishikawa, F., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED

20530913
11076861

REFERENCE
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL
MEDLINE
PUBMED

20530913
11076861

REFERENCE
AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

[illegible]

[illegible]

| | | | | | |
|-------------------|---|----|------|--|------|
| JOURNAL REFERENCE | Nature 409, 685-690 (2001) | Db | 170 | CCCCGTTGTTCTTCTACGCTTCCCGCGCGCGTGGTGGTGAAGTGGAGCGGAGCGCGAGC | 229 |
| AUTHORS | The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. | Qy | 157 | TCAGCTGCCAGATCGTGGGTAAATCCACGCCACAGGTGAGCTGGGAGAGGACCAAGCAGC | 216 |
| TITLE | Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs | Db | 230 | TCAATATGCGTGGTCTCTGGGAGAGCGCGCCGCTGCTGCTGGGAGAGGCGGCGCAGC | 289 |
| JOURNAL REFERENCE | Nature 420, 563-573 (2002) | Qy | 217 | CGGTGACGCGCGCGCGCGCTCCCTGCTGCGCCAGGACGCGACCTCTACCGCTCACA | 276 |
| AUTHORS | 6 (bases 1 to 2710) | Db | 290 | AGCTGCTGGCTCCGAGCGCTTAAGCTTTCCGAGAGGACGCGCGGAGCGGCTGCTGCG | 349 |
| | Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hasegaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, M., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Onno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, I., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. | Qy | 277 | TCCTGGAACCTGGCGCTGGGCGAGACAGTGGGCAATACCTGTGCCGCGCGCAATGCGCATAG | 336 |
| | Direct Submission | Db | 350 | TGAGCGCGCGCTGCGCCACCGACCGCGCGGTCTACGTGTGCGCGCGCGCGCAACGCGCGCG | 409 |
| | Submitted (16-JUL-2001) | Qy | 337 | GCAGAGCGCTTC-----GCTGCGGTGGGCTCCAGTGGACGCGGAGGCGCGTGGCGCG | 390 |
| | Yoshihide Hayaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216) | Db | 410 | GAGAGCGCTACGCGCGCGCGCGCTCACGCTGTGGAACCCCGCGCGCGCGCGCGAGCGCGAGC | 469 |
| | cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. | Qy | 391 | AGCAGCGCGCGCACTTCCTGCTGCGCGCCACGCTCCATCCGCTGCGCGAGGGCTCAGAGG | 450 |
| | Please visit our web site for further details. | Db | 470 | CCGAGTCCCTCCGAGTGTCCGTGCCAACACCGGCGACCGGGAGGCGCGCCCGAAGTTC | 529 |
| | URL:http://genome.gsc.riken.go.jp/ | Qy | 451 | CCACTTCCGCTGCCCGTGGTGGTCCCGAGGCGCGCAGTGGTGGTCAAGGAGC | 510 |
| | Location/Qualifiers | Db | 530 | TGACGCGCGCGCCCAATCCAGTGGGTGCTGCAAGGGGAAAGAGTGGTGTCTTACGTGCCAGG | 589 |
| | 1..2710 | Qy | 511 | GGCGCGCGCTGGGTGAGCCG | 566 |
| | /organism="Mus musculus" | Db | 590 | TGGGAGCGCTTCGCGAGCGCCCAAGCTGTACTGGGAGAGGATGGGATGGCTTGGACGAA | 649 |
| | /mol_type="mRNA" | Qy | 567 | GCAAGTGGCGCTCGCANTTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 626 |
| | /strain="C57BL/6J" | Db | 650 | GTATGGGACAGCAGCCACTTCAAGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 709 |
| | /db_xref="FANTOM,DB:9430048B09" | Qy | 627 | GCGGAGAACCCGCTGGCGCTGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 686 |
| | /db_xref="WGI.2339062" | Db | 710 | AGCTGACGTTGGCGATCTTGGCGGCA CGGCTGCCCGATTCGCGGGTGTACGTGTGTAC | 769 |
| | /db_xref="taxon:10090" | Qy | 687 | GCGGACACG | 746 |
| | /clone="9430048B09" | Db | 770 | GCCCGCAACGCGCACGCGCGCACGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 829 |
| | neck" | Qy | 747 | CGGCGGAGGCTATGCG | 797 |
| | /clone lib="RIKEN full-length enriched mouse cDNA library" | Db | 830 | GAGAGCG | 889 |
| | /dev_stages="12 days embryo" | Qy | 798 | GGCACGCGCACCTGCACTGACGCTGACTGAAGGCAAGCAGCGCGCGCGCGCGCGCGCGCGCG | 857 |
| | 635..1804 | Db | 890 | GCGCCCAAGAACTTCTGGGTGAACGAGGCGCAGCAGCGCAAAATTCGCTGCTACGTGATG | 949 |
| | /note="unnamed protein product; hypothetical Immunoglobulin and major histocompatibility complex domain containing protein (InterPro IPR03006, evidence: InterPro) | Qy | 858 | GCGGAGCCCAAGCG | 917 |
| | putative" | Db | 950 | GGCAAACTTGAGCG | 1009 |
| | /codon_start=1 | Qy | 918 | CGCCACGTGGTGTAGGAGCGCGCGAGAGAACTTCGTCTCAAGATCCTCTTCTGCAAG | 977 |
| | /protein_id="BAC28854.1" | Db | 1010 | CGCGCGCTCATGTACCG | 1069 |
| | /db_xref="GI:26330248" | Qy | 978 | CAGTGGAGCGCGCGCGCTTACACCTTGCACTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCG | 1037 |
| | /translation="MALDEVDVDSHFKLEPGRGASDEGASLTIRLLAARLPDSGVYVC HARNAHQAGALLQVHPRESPPDPENKPELPLKAGPNFVNEGKHAFC YVMPKPEIEWHLEGRPLPDRRLMYRDRDGFVLKVLVCAQKRGVYCAARNSA GQTLASVQLHVKPEPLRPLPDVVEGRHGIIVLECKVNSRIPTAFWFDORLLPC RKYQIEGAVRILVTHKLKADDDGYLCMRGRVTVANVTYKPLKPLKLDVL EGENAVLLVTEQAGVCGWSRDEGLPDTQSCSGHMHVLPVGTREDAGEITFSL GNSRTITLLRVKCSVNTYKRGGGGGRVASCQCLGWEGGKAGVLMGTCLR" | Db | 1070 | GCCAAAGACCGTGGGCTCTACGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 1129 |
| | Query Match 14.9%; Score 159.6; DB 11; Length 2710; | Qy | 1038 | TCTGTGCTGGTGTAGTGGCG | 1064 |
| | Best Local Similarity 49.9%; Pred. No. 5.2e-12; | Db | 1130 | GCGGTCCAGCTGCACGTGAAGAAACCC | 1156 |
| | Matches 493; Conservative 0; Mismatches 475; Indels 19; Gaps 3; | | | | |
| | 97 CGCCCCGCTTCTCACCG | | | | |

RESULT 8
CA368743
LOCUS
DEFINITION
5', mRNA sequence.
CA368743

CA368743
645003 NCCOWA 1RT Oncomorphous mykiss cDNA clone
496 bp
linear
EST 06-NOV-2002
1RT155J08_D_E04

```

VERSION      CA368743.1  GI:24680400
KEYWORDS     EST.
SOURCE       Oncorhynchus mykiss (rainbow trout)
ORGANISM     Oncorhynchus mykiss
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Actinopterygii; Neopterygii; Teleostei; Euteleostei;
JOURNAL      Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
COMMENT      1 (bases 1 to 496)
              Rexroad, C.E. and Keele, J.W.
              Sequence analysis of a rainbow trout normalized cDNA library
              Unpublished (2002)
              Contact: Rexroad CE
              USDA, ARS, National Center for Cool and Cold Water Aquaculture
              11876 Lestown Road, Kearneysville, WV 25430, USA
              Tel: 304 724 8340 X2129
              Fax: 304 725 0351
              Email: crexroad@cccw.ars.usda.gov
              Single pass sequencing. Bases called with phred v0.020425.c and
              trimmed with the aid of the trim_alt option. Vector identified by
              cross match v0.990329.
              Seq primer: AGCGATACCAATTTCACACAGGA.
              Location/Qualifiers
FEATURES             source
     1..496
     /organism="Oncorhynchus mykiss"
     /mol_type="mRNA"
     /db_xref="taxon:8022"
     /clone="1RT155J08 D.E04"
     /tissue_type="pooled"
     /lab_host="DH10B"
     /clone_lib="NCCCW 1RT"
     /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
     Library made from pooled tissue from brain, gill, liver,
     spleen, muscle, and kidney."
ORIGIN
Query Match      13.8%; Score 147.8; DB 14; Length 496;
Best Local Similarity 68.0%; Pred. No. 1.5e-10;
Matches 206; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY      69  GTCATGATCAGGCACAGTTTCAGCGGGCGCCCGCTTCTCACCAGCCGCGCAAGGCTTC 128
DB      179  GTATGACACAGACCTGTTTGAGAGGGCCCGCCAGTTTCTCAGCGGCCCAAGGCTTC 238

QY      129  GTGTGTGCGTGGGCAAGGACCGCCCTCAGCTGCAGATCGTGGTAAATCCACGCCA 188
DB      239  GCAGTGTGCGTGGGAAAGGATGCCACCTTGAGCTGCACCATCGTGGGCAACCCCAACCCCA 298

QY      189  CAGGTGAGCTGGGAGAGGACCGACCGGTGACGCGCGCGCGCTTCCGCTCTGGCC 248
DB      299  CTGATTACCTGGGAAAGGAGAACTAGCTGACCTCTGGGGCCGCTTCAAGATGGTG 358

QY      249  CAGGACGCGACCTCTACCGCTCACTATCTGACCTGCGCTGGGCGCAGATGGGCAA 308
DB      359  GATGATGTGACGCTTACCGCTCACCCTTACCATCTACGATCTAACCTTGGAGGACAGCGGCCAG 418

QY      309  TACGTGTCCCGCGCGCATGCGCATAGGAGGCGCTTCCGCTCCGTTGGGCGCTGCAAGTG 368
DB      419  TACATGTCCCGGCGCAAAACAATGTTGGGGAAGCATATGATGATGATGATGATGATGATGAT 478

QY      369  GAC 371
DB      479  GGC 481

RESULT 9
CNS01YES
LOCUS      Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
DEFINITION 218A06 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION  ALI72886
VERSION     ALI72886.1  GI:7810943
KEYWORDS    GSS; genome survey sequence.

```

```

SOURCE       Tetraodon nigroviridis
ORGANISM     Tetraodon nigroviridis
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Actinopterygii; Neopterygii; Teleostei; Euteleostei;
JOURNAL      Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
MEDLINE      Tetraodontidae; Tetraodontidae; Tetraodon.
PUBMED       20296633
AUTHORS      Roost Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
              Bernot, A., Fzames, C., Winkler, P., Brottier, P., Quetier, F.,
              Saurin, W. and Weissenbach, J.
              Estimate of human gene number provided by genome-wide analysis
              using Tetraodon nigroviridis DNA sequence
              Nat. Genet. 25 (2), 235-238 (2000)
              10835645
              20359837
              10899143
              3 (bases 1 to 752)
              Direct Submission
              Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : secre@genoscope.cns.fr
              Web : www.genoscope.cns.fr)
              This sequence is a single read and was generated as part of a large
              scale clone-end sequencing project of the Tetraodon nigroviridis
              genome. For more information, please take a look at
              http://www.genoscope.cns.fr/Tetraodon.
              Location/Qualifiers
FEATURES             source
     1..752
     /organism="Tetraodon nigroviridis"
     /mol_type="genomic DNA"
     /db_xref="taxon:99883"
     /clone="218A06"
     /clone_lib="G"
     /note="Genoscope sequence ID : COAG218BA03SP1-end :
     PUC-ori"
ORIGIN
Query Match      13.5%; Score 144.8; DB 29; Length 752;
Best Local Similarity 60.9%; Pred. No. 4e-10;
Matches 284; Conservative 1; Mismatches 178; Indels 3; Gaps 3;

QY      71  CATGATCAGCACAGTTTCAGCGGGCGCCCGCTTCTCACCAGCCGCGCAAGGCTTCGT 130
DB      232  CATGATCCACACCTATTTCGGGAGTCCCAAGTTTCTAACGC-GCCCGAAGCTTTTC 290

QY      131  GGTGTGCGTGGGCAAGGACCGCCCTCAGCTGCCAGATCGTGGTAAATCCACGCCACA 190
DB      291  GGTGTGTTGGGCAAGATGCGACTCTTAGCTGCACTGTGTGGGAAAGTCCGACCCCGCT 350

QY      191  GGTGAGCTGGGAGAGGACCGGTCAGCGCGCGCGCGGCTTCGCTCTGGGCCCA 250
DB      351  GATAACCTGGGAAAGGGAAGCTGAAGCTACGCTGTGGGGAGCATTTCAAGACGTTGGA 410

QY      251  GGACGGGACCTCTACCGCCCTCACTATCTTGAGCTGGGCGCTGGGCGACAGTGGGCAATA 310
DB      411  GGATGGAGATGTGTACCCGCTGACCATCTACGAGTGAACCTGGAAGACAGCGGTCACTA 470

QY      311  COTGTGCGCGCGCGCAATGCGATAGGAGCGCTTCGCTCCGCTGGGCTTCGAGGTGGA 370
DB      471  TATGTACAGAGCAAGAACAAAGTTGTGTGTAAGCGTACGCTCCGCTTACCTCAAGGTGGC 530

QY      371  CGCGGAGCGCGGTGCGCGGAGCAGCGCGCGCTTCCTGCTGGCGGCCACGCTCATCCG 430
DB      531  CCTGCGGACAGATGCCCCAG-AGGCCCTTGTCTTGTCTGCTCAAGCTGCGCCGCGCG 589

```

QY 431 CGTGGCGAGGGCTCAGAGGCCACCTTCGCTCCGGTGGGTGGCTCCCGAGGCGCGC 490
Db
590 TTAGGCTGGAGGAGAGCTTCTTCGTCGAGTGCAGCTCACCCGCGCCAC 649
QY 491 AGTAGCTGTCCAGAGCGCGCGCGCTGGGTGAGCCGACGCG 536
Db 650 TTGGA-ATGGAGAGGAGCGACGCTACCTAGGGGAGGCAATGCG 694

RESULT 10
AL927364/c
LOCUS AL927364 PJR-Z1-Z2 Danio rerio cDNA clone 188-F02-2, mRNA sequence.
DEFINITION AL927364.1 GI:23197401
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Danio rerio (zebrafish)

REFERENCE
AUTHORS
TITLE
15000 unique zebrafish EST clusters and their future use in
microarray for profiling gene expression patterns during
embryogenesis
Genome Res. 13 (3), 455-456 (2003)
22505427
12618376
Contact: Peng J
Lab of Functional Genomics
Institute of Molecular and Cell Biology
30 Medical Drive, Singapore, 117609, Singapore
Email: pengj@imcb.a-star.edu.sg
Clone requests: pengj@imcb.a-star.edu.sg.

FEATURES
source
1..539
/organism="Danio rerio"
/mol_type="mRNA"
/strain="local wildtype"
/db_xref="taxon:7955"
/clone="188-F02-2"
/tissue_type="whole embryo or fish"
/dev_stage="mixed stages"
/clone_lib="PJR-Z1+Z2"

ORIGIN
Query Match 11.9%; Score 127.4; DB 9; Length 539;
Best Local Similarity 60.1%; Pred. No. 8.3e-08;
Matches 212; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 17 ACATCCGCGCTGCCAGAGCCCCCATAGAGAGGTCCCGCCGCCACCGTCATGGA 76
Db 470 AAACATCTGCACAGTTAATATCACTATATTTGAGGGAGAAAGAGCAATGGA 411
QY 77 TCAGCCACAGTTACGCGGGCGCGCTTCTCAACCGCGCCCAAGGCTTCGTGGTGC 136
Db 410 CCAATCTATTGGCGAGCTCTCGTTTCTTCTGACCCCAAGGCTTCTCTTTGTG 351
QY 137 GGTGGGAAGACGCCACCTCAGCTGCCAGATCGTGGTAAATCCCAAGCCACAGGTGAG 196
Db 350 TGTGGAGCGGACGCGCTCTCTTAGCTGCACCTATTGTGGAAACCCCTGTACTGTAGTCAC 291
QY 197 CTGGGAGAGACACGACGCGGTGACGCGCGCGCGCTTCGTCTGGCCAGGACGG 256
Db 290 CTGGGAGAGAAAGAAATTCGTATTTCTGCGGGGGCGGATTAACAACTGTGAGGATGG 231
QY 257 CGACCTCTACCGCTCATTCTCGGACCTCGGCGCTGGCGACAGTGGGCAATACGTGTG 316
Db 230 TGACATTTATGACTTACTATCTATGATCTAACCGCTAGAGACAGTGGTCAATACATGTG 171
QY 317 CCGCGCGCGCAATAGCGAGGCGCTTCGCTCGCGCTGGGCGCTCGAGGTGG 369

Db 170 TCGTGCCAAAGACCGTTGGAGAGGCATATGCAGCTGTAAACCTGAAGGTGG 118
RESULT 11
BM799726
LOCUS BM799726
DEFINITION BM799726.1 GI:19116549
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12122 row: p column: 11
High quality sequence stop: 635.

FEATURES
source
1..1136
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5495674"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site: 1: Not;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

ORIGIN
Query Match 11.6%; Score 124.4; DB 12; Length 1136;
Best Local Similarity 59.0%; Pred. No. 2.4e-07;
Matches 237; Conservative 0; Mismatches 156; Indels 9; Gaps 1;
QY 672 GTGACTCGAGCGCGGACACGCGCCAGCCCGCGGACCTCCACGCGCGGTCTCTG 731
Db 125 GTCTAGTGTGCACCGCGCAACGCGCACGCGCCACGCGCGGGGGCGGTCTCCAG 184
QY 732 GCGCACCCTGACGCGGCGCGGAGGCTATGCGCGCGGCGCGCGCGCGCTC----- 784
Db 185 GTGCACCGCGCGCGGAGAGCGCGCGCGGAGCGCGCGCGCGCGCGCGGTGGTG 244
QY 785 --ACGCGCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 842
Db 245 GAGCCGCTCAAGTGGCGCGCTTAAGACCTTCTGGGTGACGAGGGCAAGCACCGCAAGTTC 304
QY 843 AGCTGTACTGTACCGCTG 902
Db 305 CGCTGTACTGTATGGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 364
QY 903 GTGACCGAGGGCTC 962
Db 365 CTGCTCCGCTC 424
QY 963 ATCTCTTCTGCAAGCAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1022
Db 425 GTGCTTACTGCGAGCGCAAGGATGTGGGCTTCTAGCTCTGCGCGCGCGCGCGCGCG 484

| ORIGIN | Query Match | 11.4%; | Score 122.2; | DB 13; | Length 929; |
|--------|-----------------------|---|--------------------|-----------|-------------|
| | Best Local Similarity | 61.4%; | Pred. No. 4.5e-07; | | |
| | Matches 218; | Conservative 0; | Mismatches 128; | Indels 9; | Gaps 1; |
| QY | 719 | GGCGGGCTCTGGCGCACCTGCAGGCGCGCGCGAGGTTATGCGGCCCGAGGCGCCCCC | 778 | | |
| Db | 9 | GGCGTGCTCCAGTGCACCCAGCCCGCGAGAGCCCGCGCGACCCCGACGAGGCCCC | 68 | | |
| QY | 779 | CGCCTC-----ACGCCGACGACCGGCGACCGCGACCTGCACGGTGACTGAAGCGAA | 829 | | |
| Db | 69 | CGCGCCGGTGGAGCCCGCTCAAGTGGCGGCCCTTAGACCTCTTGGTGAAACGAGGGCAA | 128 | | |
| QY | 830 | GCACGCGCGCCTCAGCTGCTACGTGACCGCGGAGCCCAAGCCCGAGACGGTGTGGAGAA | 889 | | |
| Db | 129 | GCACGCCAAAGTTCCGCTGTCTACGTGATGGCAAGCCCGAGCCCGAGATCGAATGGCACTG | 188 | | |
| QY | 890 | GGAGGCCGAGCTGGTGACCGAGGGCGCGGCCACGTGGTGTTACGAGGACCGCGAGAGAA | 949 | | |
| Db | 189 | GGAGGGCGCGCCGCTGCTCCCGAGACCGCCCGCGCTCTATGTACCGGACCGCGACGGCGG | 248 | | |
| QY | 950 | CTTGTGCTCAAGATCCTCTTTGTGCAAGCAGTCCGACCGCGGCTCTACACCTGCACGGC | 1009 | | |

| | | | | | | | |
|-----------------------|--------|--|----------|------------|------|--------|-------|
| Query Match | 11.4%; | Score | 122.2; | DB | 13; | Length | 1017; |
| Best Local Similarity | 61.4%; | Pred. No. | 4.6e-07; | | | | |
| Matches | 218; | Conservative | 0; | Mismatches | 128; | Indels | 9; |
| Gaps | | | | | | | |
| QY | 719 | GGCGCGGCTCTTGGCGCACCTGTGAGCGGGGGCGGAGGCTATGCGGCGCGCGAGGGCGCCCC | 778 | | | | |
| Db | 9 | GGCGTGCTTCAGGTGTCAACAGCCCCCCGAGAGCCCGCGCGGACCCCGACGAGGCCCC | 68 | | | | |
| QY | 779 | CGGCTC-----ACCGCCGACGACCGGACCGCCACCTGCACGGTGACTGTGAAGGC | 829 | | | | |
| Db | 69 | CGCGCGGTGGTGAGCCGCTCAAGTGC | 128 | | | | |
| QY | 830 | GCACGCGCGCCTCAGCTGTACGTGACCGCGGAGCCCAAGCCGAGACGCTGTGGAAGAA | 889 | | | | |
| Db | 129 | GCACGCCAAGTTCCGCTGCTACGTGTATGGCAAGCCGAGCCGAGATCGAATGGCACTG | 188 | | | | |
| QY | 890 | GGACGGCCAGCTGGTGACCCGAGGGCCCGGCCACGTGTGTACGAGGACCGCGCAGAGAA | 949 | | | | |
| Db | 189 | GGAGGGCCCGCCGCTGCTCCCGGACCGCGCCGCTCATGTACCGCGACCGGACCGCGG | 248 | | | | |

OM nucleic - nucleic search, using sw model
Run on: September 19, 2004, 14:16:08 ; Search time 3959.71 Seconds
(without alignments)
10628.572 Million cell updates/sec
Title: US-10-077-130-4_COPY_23150_24120
Perfect score: 971
Sequence: 1 cagcggcgccgctggccg.....ggcgcgtacaaaagtctaga 971
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544
Minimum DB seq length: 0
Maximum Match 100%
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ets.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_to.*
27: em_ats.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|------------|--------------------|
| 1 | 946.4 | 97.5 | 4175 | 6 | AX039409 | AX039409 Sequence |
| 2 | 926.8 | 95.4 | 5207 | 6 | AR255867 | AR255867 Sequence |
| 3 | 926.8 | 95.4 | 5207 | 6 | AX430856 | AX430856 Sequence |
| 4 | 926.8 | 95.4 | 5207 | 6 | AX430858 | AX430858 Sequence |
| 5 | 911.4 | 93.9 | 3225 | 6 | AX642969 | AX642969 Sequence |
| 6 | 911.4 | 93.9 | 5454 | 6 | AX504255 | AX504255 Sequence |
| 7 | 906.4 | 93.3 | 4936 | 6 | AX207401 | AX207401 Sequence |
| 8 | 905.4 | 93.2 | 5007 | 6 | AX039410 | AX039410 Sequence |
| 9 | 905.4 | 93.2 | 7928 | 6 | AX039412 | AX039412 Sequence |
| 10 | 884.4 | 91.1 | 4041 | 9 | AB046859 | AB046859 Homo sapi |
| 11 | 631.8 | 65.1 | 1183 | 10 | BC046431 | BC046431 Mus muscu |
| 12 | 262 | 27.0 | 62164 | 9 | AL670729 | AL670729 Human DNA |
| C 13 | 229.4 | 23.6 | 164766 | 2 | AC026657 | AC026657 Homo sapi |
| C 14 | 229.4 | 23.6 | 174612 | 2 | AC023889 | AC023889 Homo sapi |
| 15 | 223.4 | 23.0 | 225 | 6 | AX309418 | AX309418 Sequence |
| 16 | 207 | 21.3 | 174612 | 2 | AC023889 | AC023889 Homo sapi |
| 17 | 206 | 21.2 | 7710 | 6 | AX056397 | AX056397 Sequence |
| 18 | 204.4 | 21.1 | 7789 | 6 | AX746187 | AX746187 Sequence |
| 19 | 198.8 | 20.5 | 4357 | 10 | AK122488 | AK122488 Mus muscu |
| 20 | 189.4 | 19.5 | 10801 | 10 | AF215896 | AF215896 Mus muscu |
| 21 | 179.2 | 18.5 | 6726 | 9 | AB037718 | AB037718 Homo sapi |
| C 22 | 176.6 | 18.2 | 260998 | 2 | AC099089 | AC099089 Rattus no |
| C 23 | 173.6 | 17.9 | 103129 | 10 | AL662809 | AL662809 Mouse DNA |
| C 24 | 173.6 | 17.9 | 258622 | 2 | AC020877 | AC020877 Mus muscu |
| 25 | 126.8 | 13.1 | 182 | 6 | AX913866 | AX913866 Sequence |
| 26 | 126.8 | 13.1 | 182 | 6 | BD049399 | BD049399 Sequence |
| 27 | 113.8 | 11.7 | 1365 | 12 | AY335738 | AY335738 Synthetic |
| 28 | 113.8 | 11.7 | 2055 | 9 | AB022341 | AB022341 Homo sapi |
| 29 | 113.8 | 11.7 | 2079 | 6 | BD127287 | BD127287 Primer fo |
| 30 | 113.8 | 11.7 | 2079 | 9 | AK074799 | AK074799 Homo sapi |
| 31 | 113.8 | 11.7 | 2118 | 9 | AK097643 | AK097643 Homo sapi |
| 32 | 113.8 | 11.7 | 2224 | 6 | AX880020 | AX880020 Sequence |
| 33 | 113.8 | 11.7 | 2224 | 6 | BD012208 | BD012208 Novel gen |
| 34 | 113.8 | 11.7 | 2224 | 6 | BD158150 | BD158150 Primer fo |
| 35 | 113.8 | 11.7 | 2224 | 9 | AK027590 | AK027590 Homo sapi |
| 36 | 112.2 | 11.6 | 2105 | 6 | BD176607 | BD176607 Method of |
| 37 | 112.2 | 11.6 | 2105 | 9 | AB007144 | AB007144 Homo sapi |
| 38 | 112.2 | 11.6 | 2132 | 6 | AR076189 | AR076189 Sequence |
| 39 | 112.2 | 11.6 | 2132 | 6 | AR124102 | AR124102 Sequence |
| 40 | 112.2 | 11.6 | 2132 | 6 | E23384 | E23384 DNA encodin |
| C 41 | 105.8 | 10.9 | 1719 | 8 | AK109754 | AK109754 Oryza sat |
| 42 | 99.2 | 10.2 | 2731 | 8 | AK110172 | AK110172 Oryza sat |
| 43 | 97 | 10.0 | 3870 | 5 | CHESKELMUS | M81787 Gallus sat |
| 44 | 96 | 9.9 | 1560 | 10 | BC062076 | BC062076 Rattus no |
| 45 | 96 | 9.9 | 2903 | 4 | RAEMLCKA | J05194 Rabbit myos |

ALIGNMENTS

RESULT 1
AX039409
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AX039409
Sequence 2 from Patent WO0063381.
AX039409
AX039409.1 GI:11229476
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Zeng W., Stanton L. and Kong H.
Mammalian protein with putative function in signal transduction
Patent: WO 0063381-A 2 26-Oct-2000;
linear DNA 4175 bp PAT 18-NOV-2000

```
SCIOS INC. (US)
Location/Qualifiers
1. 4175
/organism="Homo sapiens"
/mol_type="unass.gned DNA"
/db_xref="taxon:9606"
<1. .4056
/feature="unnamed protein product"
/codon_start=1
/protein_id="CAC16624.1"
/db_xref="GI:11229477"
/db_xref="REMTREMBL:CAC16624"
/translation="GVQFSQIGSPFVSPHIIQNPVSEASDIWAGVIVSYLSLTCS
PFAGESRATLNTLEGRVSWSPMAAHLSEDAKDFIKATLQAPAPPSAAQCLSHP
WFLKSMFAEEAHFINTKQLFLLARWORSLSYKSLVWRSIPELLRPDPSPSLG
VARILCRDTGSSSSSSSSONELAPFAKSLPSPVTHSPHLLHPRGFLRPSASLPEE
AEASERTEAPAPSPGEGPAAQOCVPRHSVIRSLFYHQAEGSEPHGALAPGRR
HPARRHLKGYIAGALPGRLRPLMEHVRVLEEAAREEOATLLAKAPSPETALRLPA
SGTHLAFHSHSLSDSPSPRPSSEAGCAQLPSPAGGAPIRDMGHQPGSKQLPS
TGGHPGTAQPERPSPFPGWQAPFCHPKQASAPFQEGSPHPAVAPCPGSPFPFGCK
EAPLVPSFPGQAPAPAPAKASPLDLSKMGPDLSLPGRPQPCSPGSGASQSS
SQAFLVSSQVQTEPPGSLDARGMTQEAEDLSDSFTLQRPQEOATWKFSLGARG
SYGSLVAGVGTAFAGGAGMLGQGPWARIATAVQSSEEEQEAESAQSEBQBAR
AEASPLVQARVPVEGRAPTRSPPTWEDIGQVSLVOIRDLGDAEADTISLDI
SEVDPAVNLNLDLYDKVLPTEFMI FRKVPKSAQPERPPSPWAEELAEFPPTWPG
ELGHAGLEITESESDVALLAEAVGRKKNSSPSRSLFHFGRHLPLDEPAELGR
ERVKASVEHSIRILKRGPEGLEKEGPPKPKPLASPLSLGKSWDRAPTLRELSDET
VVLQGSVTLCQVQAQAQATWSDGAPLESSESRVLSATLXNFQTLTLVVAEDL
GVYTCVSNALGVTITGVLKRAERPSSSPDITGEVYADGVLLVWKPVESEVFTYI
VOCLEGSSTWTLASDIPDCYLTSLKSGRTYTFRTACVSKAGMGPYSPSROVLLG
GPHLASEESQSGSAOLPSTKTFATQTOIORGRFSVVRQKWEKASGRALAAK11PY
HPKDTAVLRVEYALGLRHPHLAQLHAYLSPHLYLLELCSGPELLPLAERASY
SESEVDKQWLSATQYLNQHLHLDASENNIITEYNLLKVDVLDGNQASLSQKXY
LPSDKFDYLETMAPELLEGGQVFPDIDWAGVITAFILMSAEYFVPSSEBARDLQRL
RKGLVRLSRVAGLSGGAVFLRSTLCAQWPGRPCASSCLQCPMLTEEGPACSRPAPV
TFPTARLVFVRNEKRRALLYKHNLAQVR"
ORIGIN
Query Match 97.5%; Score 946.4; DB 6; Length 4175;
Best Local Similarity 99.9%; Pred. No. 2.1e-140;
Matches 947; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAGCGGGGGCGGTGGCGGCAAGATCATCCCCCTACACCCCAAGGACAGACAGT 60
DB 3228 CAGCGGGGGCGGTGGCGGCAAGATCATCCCCCTACACCCCAAGGACAGACAGT 3287
QY 61 GCTGGCGGAATACAGAGCCCTCAGGGGCTCGGCACCCGACCTGCCCCAGCTGCACGC 120
DB 3288 GCTGGCGGAATACAGAGCCCTCAGGGGCTCGGCACCCGACCTGCCCCAGCTGCACGC 3347
QY 121 AGCTTACCTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTGCTCTGGGCGGAGCT 180
DB 3348 AGCTTACCTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTGCTCTGGGCGGAGCT 3407
QY 181 GCTCCCTGCTGGCGGAGAGGGCTCTCTACTAGAAATCCGAGTGAAGGACTACCTGTG 240
DB 3408 GCTCCCTGCTGGCGGAGAGGGCTCTCTACTAGAAATCCGAGTGAAGGACTACCTGTG 3467
QY 241 GCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGCACATCTCTGACCTGCACCTGAG 300
DB 3468 GCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGCACATCTCTGACCTGCACCTGAG 3527
QY 301 GTCCGAGAACATGATCATCAACGGAATACACCTGCTCAAGGTCTGAGACCTGGGCAATGC 360
DB 3528 GTCCGAGAACATGATCATCAACGGAATACACCTGCTCAAGGTCTGAGACCTGGGCAATGC 3587
QY 361 ACAGAGCCTCAGCAGGAGAGGTGCTGCTCCTCAGACAGATTTCAAGGACTACCTAGAGAC 420
DB 3588 ACAGAGCCTCAGCAGGAGAGGTGCTGCTCCTCAGACAGATTTCAAGGACTACCTAGAGAC 3647
QY 421 CATGGCTCCAGAGCTCTCTGGAGGCCACAGGGGCTGTTCACACAGACAGACATCTGGGCCAT 480
DB 3648 CATGGCTCCAGAGCTCTCTGGAGGCCACAGGGGCTGTTCACACAGACAGACATCTGGGCCAT 3707
```

ORIGIN

RESULT 4
AX430858

LOCUS AX430858 5207 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 3 from Patent WO0240683.
ACCESSION AX430858
VERSION AX430858.1 GI:21655922
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Ketchum, K., Beasley, E.M., Wei, M.H. and di Francesco, V.
TITLE Isolated human kinase proteins, nucleic acid molecules encoding human kinase proteins, and uses thereof
JOURNAL Patent: WO 0240683-A 3 23-MAY-2002;
PE CORP NY (US)
FEATURES
source Location/Qualifiers
1..5207
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 95.4%; Score 926.8; DB 6; Length 5207;
Best Local Similarity 99.8%; Pred. No. 2.6e-137;
Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGCGGGCGGCGTGGCGGCAAGATCATCCCTACCAACCCCAAGGACAGACAGT 60
DB 4278 CAGCGGGCGGCGTGGCGGCAAGATCATCCCTACCAACCCCAAGGACAGACAGT 4337
QY 61 GCTCGGGAATACAGAGCCCTCAAGGCGCTGCGCCACCCGACCTGGCCCGACGCG 120
DB 4338 GCTCGGGAATACAGAGCCCTCAAGGCGCTGCGCCACCCGACCTGGCCCGACGCG 4397
QY 121 AGCTTACTAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTCTGGCCCGGAGCT 180
DB 4398 AGCTTACTAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTCTGGCCCGGAGCT 4457
QY 181 GCTCCCTGCTGGCGGAGGCGCTCTCTAATCAGAAATCCGAGTGAAGACTACCTGTG 240
DB 4458 GCTCCCTGCTGGCGGAGGCGCTCTCTAATCAGAAATCCGAGTGAAGACTACCTGTG 4517
QY 241 GCAGATGTTAGTGGCACCAGTACCTGCACCAACAGACATCTGCACTGGACCTGAG 300
DB 4518 GCAGATGTTAGTGGCACCAGTACCTGCACCAACAGACATCTGCACTGGACCTGAG 4577
QY 301 GTCCGAGAAATGATCATCAACGAATCAACCTGCTCAAGTCTGTGACCTGGGCAATGC 360
DB 4578 GTCCGAGAAATGATCATCAACGAATCAACCTGCTCAAGTCTGTGACCTGGGCAATGC 4637
QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAGAGTTCAAGACTACCTAGAC 420
DB 4638 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAGAGTTCAAGACTACCTAGAC 4697
QY 421 CATGCTCCAGAGTCTCTGAGGGGCGCAGGGGCTGTTCCACAGACAGATCTGGGCGAT 480
DB 4698 CATGCTCCAGAGTCTCTGAGGGGCGCAGGGGCTGTTCCACAGACAGATCTGGGCGAT 4757
QY 481 CGGTGTGACGCTTTCATGCTGAGCGCCGAGTACCCGGTGAAGAGGTTGACG 540
DB 4758 CGGTGTGACGCTTTCATGCTGAGCGCCGAGTACCCGGTGAAGAGGTTGACG 4817
QY 541 CGACTCGAGAGAGACTCGCAGAGGGGCTGCTCCGCTGAGCGGCTGCTACGCGGGCT 600
DB 4818 CGACTCGAGAGAGACTCGCAGAGGGGCTGCTCCGCTGAGCGGCTGCTACGCGGGCT 4877
QY 601 GTCCGGGGCGCCGTTGCTCTTCTGCGCAGACCTCTGTGCGCCCGAGCCCTGGGCGCGCC 660
DB 4878 GTCCGGGGCGCCGTTGCTCTTCTGCGCAGACCTCTGTGCGCCCGAGCCCTGGGCGCGCC 4937
QY 661 CTGCGGCTCCAGCTGCTGAGTGGCTTACAGAGAGGCGCGCGCTGTTCTGCG 720

QY 721 GCCCGCGCGGCTGACCTTCCCTACCGCGGCTCGCGGTCTTCTGTCGCAATCGCGAGAA 780
DB 4998 GCCCGCGCGGCTGACCTTCCCTACCGCGGCTCGCGGTCTTCTGTCGCAATCGCGAGAA 5057
QY 781 GAGACGCGCGCTGCTGTACAGAGGCAACCTGGCCAGGTGCTGTGAGGTGCTGCCCGC 840
DB 5058 GAGACGCGCGCTGCTGTACAGAGGCAACCTGGCCAGGTGCTGTGAGGTGCTGCCCGC 5117
QY 841 GCCACACCTTGGTCTCCCGCTGGGCTGCTGACAGCGCGCAATAAAGGCGCGCGC 900
DB 5118 GCCACACCTTGGTCTCCCGCTGGGCTGCTGACAGCGCGCGCAATAAAGGCGCGCGC 5177
QY 901 CGGCGGAGAAAAAAGGCG 930
DB 5178 CGGCGGAGAAAAAAGGCG 5207
RESULT 5
AX642969 3225 bp DNA linear PAT 24-FEB-2003
LOCUS Sequence 46 from Patent WO01096547.
DEFINITION AX642969
ACCESSION AX642969
VERSION AX642969.1 GI:28550118
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
Yue, H., Lal, P., Bandman, O., Borovsky, M., Au-Young, J., Lu, Y., Gandhi, A.R., Tribouley, C.M., Walla, N., Yao, M.G., Lu, D.A., Greenwald, S.R., Ramkumar, J., Griffin, J.A., Kearney, L., Burford, N., Nguyen, D.B., Tang, Y.T., Baughn, M.R., He, A., Thornton, M., Hafalia, A., Patterson, C., Gururajan, R., Lo, T.P., Khan, F., Recipon, S.A., Azimzai, Y., Policky, J.L., Ding, L., Grether, M., Elliott, V.S., Thangavelu, K., Batra, S. and Ison, C.H.
Human kinases
TITLE Patent: WO 01096547-A 46 20-DEC-2001;
JOURNAL Incyte Genomics, Inc. (US)
FEATURES Location/Qualifiers
source 1..3225
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 3016969CB1"
ORIGIN
Query Match 93.9%; Score 911.4; DB 6; Length 3225;
Best Local Similarity 99.9%; Pred. No. 7.5e-135;
Matches 912; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAGCGGGCGGCGTGGCGGCAAGATCATCCCTACCAACCCCAAGGACAGACAGT 60
DB 2313 CAGCGGGCGGCGTGGCGGCAAGATCATCCCTACCAACCCCAAGGACAGACAGT 2372
QY 61 GCTCGCGGAATACAGAGGCGCTCAAGGCGCTGCGCCACCCGACCTGGCCCGACGCGC 120
DB 2373 GCTCGCGGAATACAGAGGCGCTCAAGGCGCTGCGCCACCCGACCTGGCCCGACGCGC 2432
QY 121 AGCTTACTAGCCCCCGGACCTGGTGTCTATCTTGGAGCTGTCTTGGGCGGCGCT 180
DB 2433 AGCTTACTAGCCCCCGGACCTGGTGTCTATCTTGGAGCTGTCTTGGGCGGCGCT 2492
QY 181 GCTCCCTGCTGGCGGAGAGGCGCTCTACTCAGAATCCGAGTGAAGGACTACCTGTG 240
DB 2493 GCTCCCTGCTGGCGGAGAGGCGCTCTACTCAGAATCCGAGTGAAGGACTACCTGTG 2552
QY 241 GCAGATGTTAGTGGACACCCAGTACTGACAAACAGACATCTCTGACCTGGACCTGAG 300
DB 2553 GCAGATGTTAGTGGACACCCAGTACTGACAAACAGACATCTCTGACCTGGACCTGAG 2612
QY 301 GTCGCGCAATCATCATCAGGAAATCAACCTGCTCAAGTCTGTGAGGCGCAATGC 360

| | | | | | | | | |
|----|------|-----------|----------------|------------|-----------------|-----------------|-------------------|------|
| Db | 2613 | GTCCGAGAA | CATGATCATCATCA | CCGAATCA | CAACCTGCTCAAGGT | CGTGAGCATCTGGC | CAATGC | 2572 |
| Qy | 361 | ACAGAGCCT | CAGCCAGAGAGAGT | TGCTGC | CCCTCAGACAAGT | TTCAGAGCAT | TACCTAGAGAC | 420 |
| Db | 2673 | ACAGAGCCT | CAGCCAGAGAGAGT | TGCTGC | CCCTCAGACAAGT | TTCAGAGCAT | TACCTAGAGAC | 2732 |
| Qy | 421 | CATGGCTC | CAGAGCTCTCTG | AGGGCCAG | GGGGTGTTCC | CACAGACAGACA | TCTGGGCCAT | 480 |
| Db | 2733 | CATGGCTC | CAGAGCTCTCTG | AGGGCCAG | GGGGTGTTCC | CACAGACAGACA | TCTGGGCCAT | 2792 |
| Qy | 481 | CGGTGTG | CACAGCCTTTC | ATCATGTG | AGCGCCGAGT | ATCCGGT | GAGCAGCAGGGTGCACG | 540 |
| Db | 2793 | CGGTGTG | CACAGCCTTTC | ATCATGTG | AGCGCCGAGT | ATCCGGT | GAGCAGCAGGGTGCACG | 2852 |
| Qy | 541 | CGACTTC | CAGAGAGGACT | CGCGAAG | GGGCTGCTCC | GGCTGAGCGGCT | GTCTACCGGGGCT | 600 |
| Db | 2853 | CGACTTC | CAGAGAGGACT | CGCGAAG | GGGCTGCTCC | GGCTGAGCGGCT | GTCTACCGGGGCT | 2912 |
| Qy | 601 | GTCCGGGG | GGCGCGTGG | CCCTTCTCTG | CGACGACTCTGTG | CGCCAGCCCTTGG | GGCGCCGGCC | 660 |
| Db | 2913 | GTCCGGGG | GGCGCGTGG | CCCTTCTCTG | CGACGACTCTGTG | CGCCAGCCCTTGG | GGCGCCGGCC | 2972 |
| Qy | 661 | CTCGGGCT | CGAGCTGCCT | TCGAGTG | CCCGTGCTTAA | CAGAGAGGGCCCG | CGCTGTTCGCG | 720 |
| Db | 2973 | CTCGGGCT | CGAGCTGCCT | TCGAGTG | CCCGTGCTTAA | CAGAGAGGGCCCG | CGCTGTTCGCG | 3032 |
| Qy | 721 | GCCCGCGC | CGGTGACCTT | CTCCCTAC | CGCGCGCTTCTT | CGTGGCCAATCG | CGCAGAA | 780 |
| Db | 3033 | GCCCGCGC | CGGTGACCTT | CTCCCTAC | CGCGCGCTTCTT | CGTGGCCAATCG | CGCAGAA | 3092 |
| Qy | 781 | GAGACGGC | CGCTGTGTAT | CAAGAGG | CAACAACCTT | GGCCCAAGGTGCG | CTGAGGGTGC | 840 |
| Db | 3093 | GAGACGGC | CGCTGTGTAT | CAAGAGG | CAACAACCTT | GGCCCAAGGTGCG | CTGAGGGTGC | 3152 |
| Qy | 841 | GCCACACCT | TGTGTCTCC | CGCTGGG | GGGTGCGTGC | AGCGGCCAATAAAAA | CGCCCAGC | 900 |
| Db | 3153 | GCCACACCT | TGTGTCTCC | CGCTGGG | GGGTGCGTGC | AGCGGCCAATAAAAA | CGCACAGC | 3212 |
| Qy | 901 | CGGGCGAG | AAAAA | 913 | | | | |
| Db | 3213 | CGGGCGAG | AAAAA | 3225 | | | | |

RESULT 6
 AX504255
 LOCUS
 DEFINITION
 Sequence 44 from Patent WO0233099.
 AX504255
 ACCESSION
 AX504255.1 GI:23386095
 KEYWORDS
 Homo sapiens (human)
 SOURCE
 Homo sapiens
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 REFERENCE
 AUTHORS
 Gururajan,R., Baughn,M.R., Wallia,N.K., Elliott,V.S., Xu,Y.,
 Arvizu,C., Yao,M.G., Ramkumar,J., Ding,L., Tang,Y.T., Hafalia,A.J.,
 Nguyen,D.B., Gandhi,A.R., Lu,Y., Yue,H., Burford,N., Bandman,O.,
 Tribouley,C.M., Lal,P.G., Recipon,S.A., Lu,D.A., Borowsky,M.L.,
 Thornton,M., Swarnaker,A., Thangavelu,K., Khan,F.A. and Ison,C.H.
 Human Kinases
 TITLE
 Patent: WO 0233099-A 44 25-APR-2002;
 INCYTE Genomics, Inc. (US)
 FEATURES
 Location/Qualifiers
 1..5454
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 7638121C1B1"

Query Match 93.9%; Score 911.4; DB 6; Length 5454;

| | | | | | | |
|----------|------------|----------|----------------|-----------|--------|-----------------|
| AX207401 | LOCUS | AX207401 | 4936 bp | DNA | linear | PAT 30-AUG-2001 |
| | DEFINITION | Sequence | 14 from Patent | WO015356. | | |
| | ACCESSION | AX207401 | | | | |

| Best Local Similarity 99.9%; Pred. No. 7e-135; | | Matches 912; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | |
|--|------|--|------|
| QY | 1 | CAGCGGGCGGGCGCTGGCGGCCCAAGATCATCCCTACCAACCACCAAGGACAAAGACAGCAGT | 60 |
| Db | 4542 | CAGCGGGCGGGCGCTGGCGGCCCAAGATCATCCCTACCAACCACCAAGGACAAAGACAGCAGT | 4601 |
| QY | 61 | GCTGCGCGAATACGAGGGCCCTCAAGGGCCCTGCGCCACCCGCACTCTGGGCCACAGTGCACGC | 120 |
| Db | 4602 | GCTGCGCGAATACGAGGGCCCTCAAGGGCCCTGCGCCACCCGCACTCTGGGCCACAGTGCACGC | 4661 |
| QY | 121 | AGCCTACCTCAGCCCCCGGCACCTGGTGTCTATCTTGGAGCTGTGTCTGGGCCCGCAGCT | 180 |
| Db | 4662 | AGCCTACCTCAGCCCCCGGCACCTGGTGTCTATCTTGGAGCTGTGTCTGGGCCCGCAGCT | 4721 |
| QY | 181 | GCTCCCTGCTGGCCGAGAGGCCCTCTACTCAGAATCCGAGGTGAAGACTACTCTGTG | 240 |
| Db | 4722 | GCTCCCTGCTGGCCGAGAGGCCCTCTACTCAGAATCCGAGGTGAAGACTACTCTGTG | 4781 |
| QY | 241 | CGAGATGTTAGTGGCCACCCAGTACTCTCACAAACAGACATCTCTGCACCTCGACCTGGACCTGAG | 300 |
| Db | 4782 | CGAGATGTTAGTGGCCACCCAGTACTCTCACAAACAGACATCTCTGCACCTGGACCTGAG | 4841 |
| QY | 301 | GTCCGAGAAACATGATCATCACCCGAATACAACCTGCTCAAGTCTGTGGACTGTGGGCAATGC | 360 |
| Db | 4842 | GTCCGAGAAACATGATCATCACCCGAATACAACCTGCTCAAGTCTGTGGACTGTGGGCAATGC | 4901 |
| QY | 361 | ACAGAGCTCAGCCAGGAGAGAGGTGCTGCCCTCAGACAACTTCAGGACTACTCTAGAGAC | 420 |
| Db | 4902 | ACAGAGCTCAGCCAGGAGAGAGGTGCTGCCCTCAGACAACTTCAGGACTACTCTAGAGAC | 4961 |
| QY | 421 | CATGCTCTCAGAGCTCCTTGAGGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGGCCAT | 480 |
| Db | 4962 | CATGCTCTCAGAGCTCCTTGAGGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGGCCAT | 5021 |
| QY | 481 | CGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCCGGTGTAGCAGCAGGGTGCACG | 540 |
| Db | 5022 | CGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCCGGTGTAGCAGCAGGGTGCACG | 5081 |
| QY | 541 | CGACTGTGACAGAGAGACTGCGCAAGGGGTGTGCGGCTGAGCCGCTGTACCGCGGGCT | 600 |
| Db | 5082 | CGACTGTGACAGAGAGACTGCGCAAGGGGTGTGCGGCTGAGCCGCTGTACCGCGGGCT | 5141 |
| QY | 601 | GTCCGGGGCGCGGTGGCGTTCTTCGCGACGACTCTGTGCGGCCAGCCTTGCGGGCCGGCC | 660 |
| Db | 5142 | GTCCGGGGCGCGGTGGCGTTCTTCGCGACGACTCTGTGCGGCCAGCCTTGCGGGCCGGCC | 5201 |
| QY | 661 | CTGCGGTCCAGCTCCCTGCACTGCGCGCTGCTTAAAGAGGAGGGCCCGGCTGTTTCGG | 720 |
| Db | 5202 | CTGCGGTCCAGCTCCCTGCACTGCGCGCTGCTTAAAGAGGAGGGCCCGGCTGTTTCGG | 5261 |
| QY | 721 | GCCCGCGCCGCTGAGCTTCCCTTACCGCGCGGCTGCGCGTCTTCTGTGCGCAATCGCGAGAA | 780 |
| Db | 5262 | GCCCGCGCCGCTGAGCTTCCCTTACCGCGCGGCTGCGCGTCTTCTGTGCGCAATCGCGAGAA | 5321 |
| QY | 781 | GAGACGCGGCTGTGTACAGAGAGGACACACTGCGCCAGGTGCGCTGAGGGTTCGCGCCG | 840 |
| Db | 5322 | GAGACGCGGCTGTGTACAGAGAGGACACACTGCGCCAGGTGCGCTGAGGGTTCGCGCCG | 5381 |
| QY | 841 | GCCACACCTTGGTCTCCCGCTGGGGGTGCGTGTGACAGCGCGCAATAAAGAACGCCACAGC | 900 |
| Db | 5382 | GCCACACCTTGGTCTCCCGCTGGGGGTGCGTGTGACAGCGCGCAATAAAGAACGCCACAGC | 5441 |
| QY | 901 | CGGGCGAGAAAA 913 | |
| Db | 5442 | CGGGCGAGAAAA 5454 | |

VERSION AX207401.1 GI:15395213
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ploeman, G., Whyte, D., Manning, G., Sudarsanam, S. and Martinez, R.
TITLE Human protein kinases and protein kinase-like enzymes
JOURNAL Patent: WO 0155356-A 14 02-AUG-2001; Sugen, Inc. (US)
FEATURES
Location/Qualifiers
1..4936
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 93.3%; Score 906.4; DB 6; Length 4936;
Best Local Similarity 99.9%; Pred. No. 4.4e-134;
Matches 907; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAGCGGGCGGCTGGCGGCAAGATCATCCCTACGACCCCAAGGACAGCAGT 60
DB 4029 CAGCGGGCGGCTGGCGGCAAGATCATCCCTACGACCCCAAGGACAGCAGT 4088
QY 61 GCTCGCGGAATACAGAGCCCTCAAGGCGCTGCGCCACCCGACCTGCGCCAGCTGCACGC 120
DB 4089 GCTCGCGGAATACAGAGCCCTCAAGGCGCTGCGCCACCCGACCTGCGCCAGCTGCACGC 4148
QY 121 AGCTACTCTAGCCCCCGGACCTGGTCTCATCTTGAGCTGTGCTCTGGCGCCGAGCT 180
DB 4149 AGCTACTCTAGCCCCCGGACCTGGTCTCATCTTGAGCTGTGCTCTGGCGCCGAGCT 4208
QY 191 GCTCCCTGCTGGCGGAGAGGCTCTCTACTAGATCCGAGTGAAGACTACCTGTG 240
DB 4209 GCTCCCTGCTGGCGGAGAGGCTCTCTACTAGATCCGAGTGAAGACTACCTGTG 4268
QY 241 CGAGATGTGAGTGCCACCCAGTACCTGCAACACGACACATCTCTGACCTGGACCTGAG 300
DB 4269 CGAGATGTGAGTGCCACCCAGTACCTGCAACACGACACATCTCTGACCTGGACCTGAG 4328
QY 301 GTCCGAGAACATGATCATACCGAATACACCTGCTCAAGGTGTGACCTGGGCAATGC 360
DB 4329 GTCCGAGAACATGATCATACCGAATACACCTGCTCAAGGTGTGACCTGGGCAATGC 4388
QY 361 ACAGAGCTCAGCAGAGAGAGTGTGCTGCTCAGCAAGTTCAGAGACTACCTAGAGAC 420
DB 4399 ACAGAGCTCAGCAGAGAGAGTGTGCTGCTCAGCAAGTTCAGAGACTACCTAGAGAC 4448
QY 421 CATGCTCCAGAGCTCTGAGAGGCGCAAGGGGCTGTTCACAGACAGATCTGCGGCAT 480
DB 4449 CATGCTCCAGAGCTCTGAGAGGCGCAAGGGGCTGTTCACAGACAGATCTGCGGCAT 4508
QY 481 CGGTGTACAGCTTTCATCTGCTGAGCGCGAGTACCCGCTGAGCAGCAGGCTGCACG 540
DB 4509 CGGTGTACAGCTTTCATCTGCTGAGCGCGAGTACCCGCTGAGCAGCAGGCTGCACG 4568
QY 541 CGACTCGAGAGAGACTCGCAAGGGGCTGTGCTGCTGAGCGGCTGTGCTACCGGGGCT 600
DB 4569 CGACTCGAGAGAGACTCGCAAGGGGCTGTGCTGCTGAGCGGCTGTGCTACCGGGGCT 4628
QY 601 GTCCGGGGCGCTGCTGCTTCTTCTGCGACACTCTGCTGCGCCGACGCTGGGGCGGCC 660
DB 4629 GTCCGGGGCGCTGCTGCTTCTTCTGCGACACTCTGCTGCGCCGACGCTGGGGCGGCC 4688
QY 661 CTGCGGCTTCAGCTGCTGAGTCCCGTGGCTTAACAGAGAGGGGCGGCTGTGCTGCGG 720
DB 4689 CTGCGGCTTCAGCTGCTGAGTCCCGTGGCTTAACAGAGAGGGGCGGCTGTGCTGCGG 4748
QY 721 GCGCGCGCGCTGAGCTTCTTCTGCGCGGCTGCGGCTCTTCTGCGAATCGCAGAA 780
DB 4749 GCGCGCGCGCTGAGCTTCTTCTGCGCGGCTGCGGCTCTTCTGCGAATCGCAGAA 4808

QY 781 GAGACGCGCGCTGTGTACAAGAGGCAACAACCTGCGCCAGGTGCGCTGAGGGTGCCTCCCG 840
DB 4809 GAGACGCGCGCTGTGTACAAGAGGCAACAACCTGCGCCAGGTGCGCTGAGGGTGCCTCCCG 4868
QY 841 GCACACACCTTGTCTCCCGCTGGGGTCTGCTGAGACGCGCCCAATAAACAACGCCAGC 900
DB 4869 GCACACACCTTGTCTCCCGCTGGGGTCTGCTGAGACGCGCCCAATAAACAACGCCAGC 4928
QY 901 CGGGCGAG 908
DB 4929 CGGGCGAG 4936
RESULT 8
AX039410
LOCUS
DEFINITION Sequence 3 from Patent WO063381.
ACCESSION AX039410
VERSION AX039410.1 GI:11229478
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Zeng, W., Stanton, L. and Kong, H.
TITLE Mammalian protein with putative function in signal transduction
JOURNAL Patent: WO 063381-A 3 26-OCT-2000; SCIOS INC. (US)
FEATURES
source
1..5007
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
97..4929
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC16625.1"
/db_xref="GI:11229479"
/translation="MQVTIEDVQAQTGTAQFEALIEGDPQPSVTWVKDSVQLVDSTRE
LSQEGGTYTLVLRHVASKDAGVYCLAQNTGQVLCXAEALLVLGDNFDSKQSH
RKUHSFYVEKEEIGRGVFGKRVQHKNKILCAKPIRSTRAQAKRERDILAA
LSHPLVTGLDQFETRTLLILSLCSSELLDLRYGVVTEAEVKYIQQLVEGHH
YHSHGLHLDIKFSNIMLVHAPEDIKIDFGFAQNTTFAELQFSGYGFVSPFI
IQONPVSEASDIWAGVISYLSLTCSSPAGESDRAITLVNLEGRVSWSPMAHLE
DAKDFIKATLQRAQAPSAACLSHFVFLKSMFAEAHINTKQLKFLARSWQS
LMSYKSLIWMESIPELLRGPPDSPSLGVARHLCRDTGSGSSSSSSSNELAPPARAKS
LPPVTHSPULLHPRGFLRPSASLPEAEASERSTEAPAPASPEGAGPAAGCVR
HSVIRSLYHQAGSPERHAGALAPGRHRRHLLKGYIAGALGRLPMEHRVL
EEAAREQATLLAKAFSTALRLAPSGHSLAPGSHSHUEHDSPTSPRSESCGA
QLRPSAPGGAPIRDMGHPGSKQLPSTGHGTPAQPERSPDPSPWQGPAPFCHPKQ
SAPOGCSPPHAPVPCPPGPPGCKEAPLVSPFLGQPPAPPAKAPAKSPPLDKQ
GPDISLPRKPGPCSPGSAQSSQSSSLRVGSQVTEPGTEPGSLDAEGMTQEA
DLSPSTLPORPOQATMRKESLAGRGYAGVAGYTFAGGDAGGVLGGPMWRIA
WAVQSEEEEEAARASQSEEQEARESPLPQVSARPPVPCRASTRSPETPHE
DIGVSLVQIRLSDAEADTIISIDISEVDFAIYNSLDLIYIKIIFEFMFKVFK
SAQPEPSPMAEEELAFPEFTWPGELGHAGLEITSEEDVALLAAVAAGRRK
WSSRSILHFFPGRHLPLDEALGRERKVAVEHISRLKGRPELEKEGPPRKP
GLASFRSLGKSWDRAPTFRLSDETVILQSVTLACQSAQAPAAQATSKDAPLE
SSSVLSATLNFQLLTILVVAEDLVGTCVSNALGTVTTTGLVKRKEPSSPC
PDIGVADGLVLMKVESPVTVYVQCSLEGGSWTTLASDIFDCYLTSKLSRG
TYTRTACVSKAGMPYSSPSEQVLLGAPSHLASEESQGRSAQPLSTKFAQTQI
QRFLSVVRQCKWAKSARLAAKLIPTPKDXTAVLREYALKLRHLRHLAQLHAAL
SFRHLVLIELCSGPELLCLAAERASYSSESEKDYLMQMLSATQILHNQHLHLLDLAS
ENMIITYNLKLVVDLGNASLQKVLPSDKFDKDYLETWAPLEEGGAVPQTDWA
IGVTFIMLSAEYVPSGADRLGRKGLVRLSRCYAGLSGGAFLRSTLCAQPW
GRPCASSCLQCPWLTTEGPACSRPAPVFTPTARLVRVFRNRKRALLYKRNLAQVR

ORIGIN

Query Match

93.2%; Score 905.4; DB 6; Length 5007;

| | | | | |
|--|------|---|------|--|
| Best Local Similarity 99.9%; Pred. No. 6.3e-134; Matches 906; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | | | |
| QY | 1 | CAGCGGGCGGCGCTGGCCGCAAGATCATCCCTTACACCCCAAGGACAGACAGCAGT | 60 | |
| DB | 4101 | CAGCGGGCGGCGCTGGCCGCAAGATCATCCCTTACACCCCAAGGACAGACAGCAGT | 4160 | |
| QY | 61 | GCTGCGCAATACGAGGCGCTCAAGGGCTCGGCAACCCGACCTGCGCCAGTGCACG | 120 | |
| DB | 4161 | GCTGCGCAATACGAGGCGCTCAAGGGCTCGGCAACCCGACCTGCGCCAGTGCACG | 4220 | |
| QY | 121 | AGCCTACCTCAGCCCGCCGACCTGGTGTCTATCTTGAGCTGTCTCTGGCCCGAGCT | 180 | |
| DB | 4221 | AGCCTACCTCAGCCCGCCGACCTGGTGTCTATCTTGAGCTGTCTCTGGCCCGAGCT | 4280 | |
| QY | 181 | GCTCCCTCTGCTGGCCGAGAGGGCTCTCTACTCAGAAATCCGAGGTGAAGACTACCTGTG | 240 | |
| DB | 4281 | GCTCCCTCTGCTGGCCGAGAGGGCTCTCTACTCAGAAATCCGAGGTGAAGACTACCTGTG | 4340 | |
| QY | 241 | GCAGATGTTGAGTGCACCCAGTACTTGCACAAACAGGACATCTGACCTGGACCTGAG | 300 | |
| DB | 4341 | GCAGATGTTGAGTGCACCCAGTACTTGCACAAACAGGACATCTGACCTGGACCTGAG | 4400 | |
| QY | 301 | GTCGAGAACATGATCATCACCAGATACAACTGCTCAAGGTCTGACCTGGCAATGC | 360 | |
| DB | 4401 | GTCGAGAACATGATCATCACCAGATACAACTGCTCAAGGTCTGACCTGGCAATGC | 4460 | |
| QY | 361 | ACAGAGCCTCAGCAGAGAGAGTGTCTGCTCCTCAGACAGTTCAGAGCTACCTAGAGAC | 420 | |
| DB | 4461 | ACAGAGCCTCAGCAGAGAGAGTGTCTGCTCCTCAGACAGTTCAGAGCTACCTAGAGAC | 4520 | |
| QY | 421 | CATGCTCCTCAGAGCTCTCTGAGGGGCGAGGGGCTGTTCCACAGACAGACATCTGGGCGAT | 480 | |
| DB | 4521 | CATGCTCCTCAGAGCTCTCTGAGGGGCGAGGGGCTGTTCCACAGACAGACATCTGGGCGAT | 4880 | |
| QY | 481 | CGGTGTGACAGCCTTCATCTGAGCGCGAGTACCCGTGAGCAGCGAGGTGACG | 540 | |
| DB | 4581 | CGGTGTGACAGCCTTCATCTGAGCGCGAGTACCCGTGAGCAGCGAGGTGACG | 4640 | |
| QY | 541 | CGACCTCAGAGAGAGTCTGCGAGGGGCTGCTCGGCTGAGCCGCTGCTACCGGGGCT | 600 | |
| DB | 4641 | CGACCTCAGAGAGAGTCTGCGAGGGGCTGCTCGGCTGAGCCGCTGCTACCGGGGCT | 4700 | |
| QY | 601 | GTCGGGGGCGCGCTGCTTCTGCGGAGCAGTCTGTCGCCCGCAGCCCTGGGGCGCGCC | 660 | |
| DB | 4701 | GTCGGGGGCGCGCTGCTTCTGCGGAGCAGTCTGTCGCCCGCAGCCCTGGGGCGCGCC | 4760 | |
| QY | 661 | CTGCGGCTCAGAGTCTGCGAGTGGCGGCTAACAGAGAGGGCGCGGCTGCTGCGG | 720 | |
| DB | 4761 | CTGCGGCTCAGAGTCTGCGAGTGGCGGCTAACAGAGAGGGCGCGGCTGCTGCGG | 4820 | |
| QY | 721 | GCCCGCGCGTGACCTTCCCTACCGCGCGGCTGCGGCTCTTCTGCGCAATCGCGAGAA | 780 | |
| DB | 4821 | GCCCGCGCGTGACCTTCCCTACCGCGCGGCTGCGGCTCTTCTGCGCAATCGCGAGAA | 4880 | |
| QY | 781 | GAGACCGCGCTGCTGTACAAGAGCACAACTGCGCCAGTGGCGCTGAGGGTCCGCCCG | 840 | |
| DB | 4881 | GAGACCGCGCTGCTGTACAAGAGCACAACTGCGCCAGTGGCGCTGAGGGTCCGCCCG | 4940 | |
| QY | 841 | GCCACACCTTGGTCTCCCGCTGGGGTGGCTGCGAGCGCGCCCAATAAAGACCCGAGC | 900 | |
| DB | 4941 | GCCACACCTTGGTCTCCCGCTGGGGTGGCTGCGAGCGCGCCCAATAAAGACCCGAGC | 5000 | |
| QY | 901 | CGGGCGA 907 | | |
| DB | 5001 | CGGGCGA 5007 | | |
| RESULT 9 AX039412 LOCUS DEFINITION Sequence 5 from Patent WO0063381. ACCESSION AX039412 | | | | |

| | | | | |
|------------|-------------|--|--|--|
| AX039412.1 | GI:11229480 | | | |
| VERSION | | | | |
| KEYWORDS | | | | |
| SOURCE | | | | |
| ORGANISM | | | | |
| REFERENCE | | | | |
| AUTHORS | | | | |
| TITLE | | | | |
| JOURNAL | | | | |
| FEATURES | | | | |
| source | | | | |
| CDS | | | | |
| ORIGIN | | | | |

ORIGIN

Query Match 93.2%; Score 905.4; DB 6; Length 7928;
Best Local Similarity 99.9%; Pred. No. 6e-134;
Matches 906; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 9
AX039412
LOCUS
DEFINITION Sequence 5 from Patent WO0063381.
ACCESSION AX039412


```
QY 1 CAGCGGGGGCGCTGGCGCCAGATCATCCCTACCAACCCCAAGCAAGACAGCAGT 60
Db 7022 CAGCGGGGGGGCGCTGGCGCCAGATCATCCCTACCAACCCCAAGCAAGACAGCAGT 7081

QY 61 GCTGCGGAATACAGAGGCGCTCAAGGGCTCGCGCACCCGACCTGGCCAGCTGCACGC 120
Db 7082 GCTGCGGAATACAGAGGCGCTCAAGGGCTCGCGCACCCGACCTGGCCAGCTGCACGC 7141

QY 121 AGCTTACCTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTGCTCTGGGCCCGAGCT 180
Db 7142 AGCTTACCTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTGCTCTGGGCCCGAGCT 7201

QY 181 GCTCCCTGCTGCGCGAGAGGCGCTCTCTCTCAGAAATCCGAGCTGAAGGACTACCTGTG 240
Db 7202 GCTCCCTGCTGCGCGAGAGGCGCTCTCTCTCAGAAATCCGAGCTGAAGGACTACCTGTG 7261

QY 241 GCAGATGTTGAGTGCACCCAGTACCTGCAACAACAGCAGACATCTCTGACCTGACCTGAG 300
Db 7262 GCAGATGTTGAGTGCACCCAGTACCTGCAACAACAGCAGCAGACATCTCTGACCTGAG 7321

QY 301 GTCCGAGAAATGATCATCAGCAATCAACCTCTCAAGTCTGAGCTGGGCAATGC 360
Db 7322 GTCCGAGAAATGATCATCAGCAATCAACCTCTCAAGTCTGAGCTGGGCAATGC 7381

QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGCTCCTCAGACAAAGTTCAAGGACTACCTGAGAC 420
Db 7382 ACAGAGCTCAGCCAGGAGAGGTGCTGCTCCTCAGACAAAGTTCAAGGACTACCTGAGAC 7441

QY 421 CATGCTCCAGAGCTCTGGAGGCGCAGGGGCTGTTCCACAGACACATCTGGGCCAT 480
Db 7442 CATGCTCCAGAGCTCTGGAGGCGCAGGGGCTGTTCCACAGACACATCTGGGCCAT 7501

QY 481 CGGTGTGACAGCTTCATCATGCTGAGCGCGAGTACCCGCTGAGCAGCGAGGTGCACG 540
Db 7502 CGGTGTGACAGCTTCATCATGCTGAGCGCGAGTACCCGCTGAGCAGCGAGGTGCACG 7561

QY 541 CGACTCAGAGAGGACTGCGCAAGGGGCTGGTCCGGCTGAGCGCTGCTACCGGGGCT 600
Db 7562 CGACTCAGAGAGGACTGCGCAAGGGGCTGGTCCGGCTGAGCGCTGCTACCGGGGCT 7621

QY 601 GTCCGGGGCGCGCTGGGCTTCTCTGCGCAGACACTCTGTGCGCCAGCGCTGGGGCCCGGCC 660
Db 7622 GTCCGGGGCGCGCTGGGCTTCTCTGCGCAGACACTCTGTGCGCCAGCGCTGGGGCCCGGCC 7681

QY 661 CTGCGCTCAGCTGCTGAGTGCCTGAGTGCCTTAACAGAGAGGGCGCGGCTGTTCGCG 720
Db 7682 CTGCGCTCAGCTGCTGAGTGCCTGAGTGCCTTAACAGAGAGGGCGCGGCTGTTCGCG 7741

QY 721 GCCCGCGCGCTGACCTTCCCTACCGCGGCTGCGGCTCTTCTGTCGCAATCGCAGAA 780
Db 7742 GCCCGCGCGCTGACCTTCCCTACCGCGGCTGCGGCTCTTCTGTCGCAATCGCAGAA 7801

QY 781 GAGACGCGCTGTGCTTAACAGAGCAACAACCTGGCCAGTGTGCTGAGGGTGCCTCCG 840
Db 7802 GAGACGCGCTGTGCTTAACAGAGCAACAACCTGGCCAGTGTGCTGAGGGTGCCTCCG 7861

QY 841 GCCACACCTTGTCTCCCGCTGGGGTGTGCTGAGCGCGCAATAAATAAAGCCCGAGC 900
Db 7862 GCCACACCTTGTCTCCCGCTGGGGTGTGCTGAGCGCGCAATAAATAAAGCCAGC 7921

QY 901 CGGCGCA 907
Db 7922 CGGCGCA 7928
```

```
RESULT 10
LOCUS AB046859
DEFINITION Homo sapiens mRNA for KIAA1639 protein, partial cds.
ACCESSION AB046859
VERSION AB046859.1 GI:10047354
KEYWORDS
```

```
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Nagase,T., Kikuno,R., Nakayama,M., Hirose,M. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
XVIII. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro
JOURNAL DNA Res. 7 (4), 273-281 (2000)
MEDLINE 20450683
PUBMED 10997877
REFERENCE 2 (bases 1 to 4041)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
292-0812, Japan [E-mail:cdnaif@kazusa.or.jp],
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
Fax:81-438-52-3914
FEATURES
Location/Qualifiers
1..4041
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="fj06072"
/notes="vector:pBluescriptII SK plus"
1..4041
/gene="KIAA1639"
<1..3962
/gene="KIAA1639"
/notes="start codon is not identified. fj06072 cDNA clone
for KIAA1639 has a 55-bp deletion at the position between
2845 and 2901, and a 1-bp insertion after the position
1640 of the sequence of KIAA1639, respectively."
/codon_start=3
/product="KIAA1639 protein"
/protein_id="BAB13465.1"
/db_xref="GI:10047355"
/translation="AMGVISYLSLTCSPFPAGESDRATLLNVLGRVSWSPMAHLIS
EDAKFIKATLQRAQAPSAACLSHPFLKSNPAEBAHFINTKQLFLARWDR
SIMSYKSLVNRSPIDELLRGPDSPLGVARHLCDTGGSSSSSSDNEIAPFARAK
SLPPSPVTHSLHPRGLRPSASLPEAEASERSTAPAPAPAGAGPAAQCCVP
LEEARREOATLLAKAPSFETALRLPASGTHLAPGHSHSLHSDSPSTPRSSACCE
ACRLPSRSGGAPINDMGHPQGSQQLSTGHPGTAQPERSPSPWQAPAPCHPQK
GSAPEGCCSPHAPVAPCPGSPFPGCKEAPLVFPSSFLGQFPAPAPAKASPFDK
MGPGDISLPGRKPGPCSPGSAQSSQSVLSRVSSQVTEFPGLSDAEGWTQEA
EDLSSTPTLQRPQEQATMRKXSLGGRGVAGVAGTFAFGDAGMLGGGPMWAI
AWAVSQSEEEQEARASQSEEQEAEASPLQVSGARPVEVGRAPTRSPPTPW
EDIGVSLVQIRDLSDGAEADTISLSDISEVDPAYLNLSDLYDKYLPFFPMFERKVP
KSAQEPSPNAERELAFEPRTWEPGELGPHAGLETTESEVDLALLAGVGRKX
KNSPSRSLFHPGRLPLDEPAGLRLRVKASVEHISRLKGRPGLEKEGPRKX
PCLASFRUSGLKSWDRAPTFDELSDETIVLGQSVTLACQVSAQPAQATWSKDGAPL
ESSRSVLATLKNFQLTILVVAEDLVGTVCSVSNALGTVTTGVLKRAERPSSP
CPDTEGVADVGLLWVKVESYGVTVIVQCSLEGSWTTLASDIPCCYLTSLKSG
GTYTFRTACVSKAGMGVPYSPSEVQLLGGPSHLASEESQGRSAPLPSTFPAFTQ
TORGRFVVROCKWEKASGRALAAKIIPVHPKDKTAVLREYBALKGLRPHLAQHAAY
LSRHLVLILFELCGPELLCLAAERASVSEVKDYLWOMLSATQYLHNOHLLHLRL
SENMLITYNLLKVVDLNGAQLSQEKVLPSPDKFDYLETMELLEGGGAVPQTDIMAI
GVTAFLMUSABEYVPSSEBARDLQRLKGLRVLRSCYAGLSGGAVAFIRSTLCAQPWG
RPCASSCLQCFLWTEEGPACSRPAPVPTFTARLRFVVRNRRKRLLYKRNHLAQVR"
```

ORIGIN

```
Query Match 91.1%; Score 884.4; DB 9; Length 4041;
Best Local Similarity 99.2%; Pred No. 1.4e-130;
Matches 901; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 CAGCGGGGGCGCTGGCGCCAGATCATCCCTACCAACCCCAAGCAAGACAGCAGT 60
Db 3140 CAGCGGGGGCGCTGGCGCCAGATCATCCCTACCAACCCCAAGCAAGACAGCAGT 3199

QY 61 GCTGCGCGAATACAGAGGCGCTTCAGCGCGCTGCGCCAGCTGGCCAGCTGCACGC 120
```

```
Db 3200 GCTGCGCAATACAGAGGCGCTCAAGGGCTCGCGCACCCGACCTGGCCACGCTGCACGC 3259
QY 121 AGCTTACCTCAGCCCGCCGACCTGGTGTCTCATTTGGAGTGTGTCTTGGGCGCGGAGCT 180
Db 3260 AGCTTACCTCAGCCCGCCGACCTGGTGTCTCATTTGGAGTGTGTCTTGGGCGCGGAGCT 3319
QY 181 GCTCCCTGCTTGGCGAGAGGGCTCTTCTACTCAGATCCAGGTGCGAGTGAAGACTACCTGTG 240
Db 3320 GCTCCCTGCTTGGCGAGAGGGCTCTTCTACTCAGATCCAGGTGCGAGTGAAGACTACCTGTG 3379
QY 241 GCAGATGTTAGTGTGACCCAGTACCTGACCAACAGCACATCTGCACCTGGACCTGAG 300
Db 3380 GCAGATGTTAGTGTGACCCAGTACCTGACCAACAGCACATCTGCACCTGGACCTGAG 3439
QY 301 GTCGAGAACATGATCATCCGCAATACAACTGCTCAAGTCTGTGACCTGGGCAATGC 360
Db 3440 GTCGAGAACATGATCATCCGCAATACAACTGCTCAAGTCTGTGACCTGGGCAATGC 3499
QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGCTCTCAGACAGTTCAGAGCTACCTAGAGAC 420
Db 3500 ACAGAGCTCAGCCAGGAGAGGTGCTGCTCTCAGACAGTTCAGAGCTACCTAGAGAC 3559
QY 421 CATGGCTCCAGAGCTCCTGGAGGCGCAGGGGCTGTTCCACAGACAGACATCTGGGCGAT 480
Db 3560 CATG-----GAGCTCTGGAGGCGCAGGGGCTGTTCCACAGACAGACATCTGGGCGAT 3613
QY 481 CGGTGTGACAGCTTCATCTGAGGCGCGAGTACCCGTTGAGCGAGGAGGTGACG 540
Db 3614 CGGTGTGACAGCTTCATCTGAGGCGCGAGTACCCGTTGAGCGAGGAGGTGACG 3673
QY 541 CGACTGTGAGAGAGACTGCGCAAGGGCTGCTCGGCTGAGCGGCTGCTACGCGGGCT 600
Db 3674 CGACTGTGAGAGAGACTGCGCAAGGGCTGCTCGGCTGAGCGGCTGCTACGCGGGCT 3733
QY 601 GTCGCGGGGCGCGTGGCTTCTTCTGCGAGACATCTGTGCGCCCGAGCCCTGGGGCGGGC 660
Db 3734 GTCGCGGGGCGCGTGGCTTCTTCTGCGAGACATCTGTGCGCCCGAGCCCTGGGGCGGGC 3793
QY 661 CTGGCGCTCCAGCTGCTGCTGAGTGCCTGAGTGCCTGAGTGCCTGAGAGAGGCGCGCTGTCGCG 720
Db 3794 CTGGCGCTCCAGCTGCTGCTGAGTGCCTGAGTGCCTGAGTGCCTGAGAGAGGCGCGCTGTCGCG 3853
QY 721 GCCCGCGCGCTGACTTCCCTACCGCGGCGCTGCGGCTTCTGCGCAATCGCGAGAA 780
Db 3854 GCCCGCGCGCTGACTTCCCTACCGCGGCGCTGCGGCTTCTGCGCAATCGCGAGAA 3913
QY 781 GAGAGCGCGCTGCTGTACAGAGGACACACTGCGCGAGTGCCTGAGGCTGCGCCCG 840
Db 3914 GAGAGCGCGCTGCTGTACAGAGGACACACTGCGCGAGTGCCTGAGGCTGCGCCCG 3973
QY 841 GCCACACCTTGGTCTTCCCGCTGGGGTCTGCTGACAGCGCGCAATAAAGGCGCCGAGC 900
Db 3974 GCCACACCTTGGTCTTCCCGCTGGGGTCTGCTGACAGCGCGCAATAAAGGCGCCGAGC 4033
QY 901 CGGCGGAG 908
Db 4034 CGGCGGAG 4041

RESULT 11
LOCUS BC046431 1183 bp mRNA linear ROD 16-SEP-2003
DEFINITION Mus musculus cDNA sequence BC046431, mRNA (cDNA clone IMAGE:4006702), partial cds.
ACCESSION BC046431
VERSION BC046431.1 GI:28386252
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1183)
```

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Schetz, J.E., Brownstein, M.J., Udín, F.B., Toshikiyuki, S., Carrinchi, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.C., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kesteven, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 1183)

Direct Submission

Strausberg, R.

Submitted (03-FEB-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Place: 92 Row: d Column: 18.

FEATURES

source

1..1183

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CZECH II"

/db_xref="taxon:10090"

/clone="IMAGE:4006702"

/tissue_type="Mammary tumor metastasized to lung."

MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR enhancer."

/clone_lib="NCI CGAP_Lu30"

/lab_host="DH10B"

/note="vector: PCMV-SPORT6"

<1..1183

/gene="BC046431"

/db_xref="LocusID:216790"

/db_xref="WGI:2664530"

<1..1027

/gene="BC046431"

/codon_start=2

/product="BC046431 protein"

/protein_id="AAH46431.1"

/db_xref="GI:28386253"

/db_xref="LocusID:216790"

/db_xref="WGI:2664530"

gene

CDS

```
/translation="VSKAGMGYPSSPEQVLLGGPNHLLASEEBSRRPAQLLPSTKT
FAFQMRFRGFRSVCRCREKASGRALAAKIVYQBEDKTAVLREYALKRLHHPHLA
QLHAYLSPRHLVILIELSCPELLSLAESYSSEVDKYLQWMLSATYLYLHAQHI
LHLDLRSENMMVTEYNLLKVIDLGNASQSDOEKVPAPENPKDYLETMAPELLEGGQAV
PQDIWAIGVTAFTIMLSGEYPESSEGTDLQKRLGLRLSRCYAGLSGGAVAFLOS
SLCAQPNRGRPCASTCLQCGWLTBEGPTGSRPTVPTPTVRLRAFVREKRALLYKK
HNLAQVR"
misc_feature
134..892
/gene="BC046431"
/note="S-Trk; Region: Serine/Threonine protein kinases,
catalytic domain"
/db_xref="CDD:smart00220"

ORIGIN
Query Match 65.1%; Score 631.8; DB 10; Length 1183;
Best Local Similarity 79.2%; Pred. No. 1.7e-90;
Matches 763; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

Qy 2 AGCGGGCGGGCGCTGGCGGCGCAAGATCATCCCTTACCACCCCAAGGACAAAGACAGCAGTG 61
Db 200 AGTGGCGGGCGCTAGCTGTCTAGATCGTTCCCTTACCACCTGAGGACAAAGACAGCTGTA 259
Qy 62 CTGGCGGNAACGAGGCGCTCAAGGCGCTGGCGCACCGGACCTGGCGGAGCTGCAGCA 121
Db 260 CTAAGAGAATACGAGGCGCTTAAGAGACTGCACACCCACATCTGGCGCAACTCCATGCC 319
Qy 122 GCCTACCTCAGCCCGCGCACCTGGTCTCATCTTGAGCTGTGCTCTGGCGCGGAGCTG 181
Db 320 GCCTACCTCAGTCCCGCGCACCTGGTCTCATCTTGAGCTGTGCTCTGGCGCGGAGCTG 379
Qy 182 CTCCTCTGCTGGCGGAGAGGCGCTCTACTCAGATTCGAGGTGAGGACTACCTGTGG 241
Db 380 CTACCTCTCTTGGCGGAGAGGAGTCTCTACTCAGAGTCTGATGTGAGGACTACCTGTGG 439
Qy 242 CAGATGTTGAGTGCACCCAGTACTCTGCACAAACAGACACATCTGCACTGGACCTGAGG 301
Db 440 CAGATGTTGAGTGCACCCAGTACTCTGCATGCCCAACACATCTGCACTGGACCTGAGG 499
Qy 302 TCCGAGAACATGATCATCAGCAATACAACTGCTCAAGCTGTGGACCTGGGCAATGCA 361
Db 500 TCCGAGAACATGATGTTGCTCAGTACAACTGCTTAAAGTTATAGACTGGGAAATGCC 559
Qy 362 CAGAGCTTCAGCAGGAGAGGTCTGCTCCTCAGACAGTTCAGGACTACCTAGAGACC 421
Db 560 CAGAGTCTCGACCAAGAGAGGTCCAGCCCTGAGAACTTCAAGAACTACCTGGAGACC 619
Qy 422 ATGGCTCAGAGCTCTCTGGAGGCGCAGGGGCTGTTCCACAGACAGACATCTGGGCGATC 481
Db 620 ATGGCTCAGAACTCTCTGGAGGCTCAAGGGCTGTTCCACAGACAGACATTTGGGCTATT 679
Qy 482 GGTGTACAGCTTCATCTGCTGAGCGCGAGTACCCGCTGAGCAGCAGGAGTGCAGCG 541
Db 680 GGTGTACAGCTTCATCTGCTGAGTGGGAGTACCCAGAGACAGCAGGAGGAGTCTGT 739
Qy 542 GACCTGACAGAGACTGCGCAAGGGCTGTGCTCGGCTGAGCGGCTGCTACCGGGGGCTG 601
Db 740 GACCTGACAGAGACTTGGCAAGGGACTCATTCGGTTGAGTCTGCTGCTATGACGATTAT 759
Qy 602 TCCGGGGCGCGCTGCTGCTGCGGAGCATCTGTGCGGCCACAGCCCTGGGGCGGGCCC 661
Db 800 TCAGGAGAGCGTAGGCTCTCTGAGAGTTTCAATATGCTCAACCCCTGGGGCGGGCG 859
Qy 662 TGGCGCTCCAGCTCTCTGAGTGGCTGCTTAAAGAGAGGCGCGGCTGTGCTGCGCG 721
Db 860 TGGCGCTCCAGCTCTGCAATGCGGTGCTGCTGAGAGAGGCGCGGCTGCTGCGG 919
Qy 722 CCGCGCGCGGTGACTCTTCCCTTACCGCGCGGTGCGGCTCTTCTGTGCGCAATTCGCGAGAG 781
Db 920 CCGCGCGCGGTGACTCTTCCCTTACCGCGCGGTGCGGCTCTTCTGTGCGCAATTCGCGAGAG 979
Qy 782 AGAGCGCGGTGCTGACAGAGAGACAACTGCGCGCGGTGCGGCTGAGGCTGCGGCGG 841
Db 980 CGGCGTGGCTCTCTTACAGAGAGACAACTGCGCGCGGTGCGGCTGAGGCTGCGGCGG ---TCCAGCT 1036
```

```
Qy 842 CCACACCCCTTGGTCTCCCGCTGGGGTCTGCAGACGGCCCAATAAAAACGCCAGCC 901
Db 1037 CTACAGACAGATGTGCATCGCACATCGCGGGACACCCCTTATTAAGATCAAAAC 1096
Qy 902 GGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 961
Db 1097 GGGCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 1156
Qy 962 AAA 964
Db 1157 AAA 1159

RESULT 12
AL670729
LOCUS
DEFINITION Human DNA sequence from clone RP11-245P10 on chromosome 1, complete
sequence.
ACCESSION AL670729
VERSION AL670729.19 GI:22316158
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 62164)
AUTHORS Van Hellmond, Z.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Aug 19, 2002 this sequence version replaced gi:21727388.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr1

RP11-245P10 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6

```
FEATURES
source
1..62164
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-245P10"
/clone_lib="RPCI-11.1"
```

ORIGIN

Query Match 27.0%; Score 262; DB 9; Length 62164;
Best Local Similarity 98.1%; Pred. No. 3.8e-32;
Matches 265; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 641 GCCAGCCCTGGGCGGCGCTGCGCTCCAGCTGCTGAGTCCCTGAGTCCCGTGGCTAACAGAG 700
DB |||||
QY 31970 GTCCGTCACAGGGGCGGCGCTGCGCTCCAGCTGCTGAGTCCCGTGGCTAACAGAG 32029
DB |||||
QY 701 GAGGCGCGGCGCTGCTGCGGCGGCGGCGCTGAGTCCCTGAGTCCCGTGGCTAACAGAG 760
DB |||||
QY 32030 GAGGCGCGGCGCTGCTGCGGCGGCGGCGCTGAGTCCCTGAGTCCCGTGGCTAACAGAG 32089
DB |||||
QY 761 TTGCTGCGCAATCGCGAAGAGAGCGCGCTGCTGACAGAGCGCAACCTGGCCGAG 820
DB |||||
QY 32090 TTGCTGCGCAATCGCGAAGAGAGCGCGCTGCTGACAGAGCGCAACCTGGCCGAG 32149
DB |||||
QY 821 GTGCGCTGAGGCTGCGGCGGCGGCGGCGCTGAGTCCCTGAGTCCCGTGGCTAACAGAG 880
DB |||||
QY 32150 GTGCGCTGAGGCTGCGGCGGCGGCGGCGCTGAGTCCCTGAGTCCCGTGGCTAACAGAG 32209
DB |||||
QY 881 CGCCAATAAAAAACGCGGCGGCGGAGAA 910
DB |||||
QY 32210 CGCCAATAAAAAACGCGGCGGCGGAGAA 32239
DB |||||

RESULT 13

AC026657/c
LOCUS AC026657 164766 bp DNA linear HTG 01-SEP-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-245P10, WORKING DRAFT
SEQUENCE, 31 unordered pieces.
ACCESSION AC026657
VERSION AC026657.4 GI:9958202
KEYWORDS HTG; HIGS_PHASE1; HIGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164766)
Waterston,R.H.
The sequence of Homo sapiens clone
2 (bases 1 to 164766)
Waterston,R.H.
Direct Submission
Submitted (22-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7637349.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H.NH024P10
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing method: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 139884 bases at least Q40
Consensus quality: 147686 bases at least Q30
Consensus quality: 151469 bases at least Q20
Insert size: 169000; agarose-ff
Insert size: 161074; sum-of-contigs
Quality coverage: 3.60 in Q20 bases; agarose-ff
Quality coverage: 3.92 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1124: contig of 1124 bp in length
1125: gap of unknown length
1224: contig of 1808 bp in length
1225: gap of unknown length
3032: contig of 1361 bp in length
3033: gap of unknown length
3132: contig of 1361 bp in length
3133: gap of unknown length
4494: contig of 1267 bp in length
4594: gap of unknown length
5860: contig of 1267 bp in length
5861: gap of unknown length
7670: contig of 1710 bp in length
7671: gap of unknown length
7771: contig of 1899 bp in length
9669: gap of unknown length
9670: contig of 1946 bp in length
11715: gap of unknown length
11716: contig of 2428 bp in length
14243: gap of unknown length
14244: contig of 2544 bp in length
16887: gap of unknown length
16888: contig of 2260 bp in length
19247: gap of unknown length
19248: contig of 2028 bp in length
21375: gap of unknown length
21376: contig of 3550 bp in length
25025: gap of unknown length
25026: contig of 2926 bp in length
28051: gap of unknown length
28052: contig of 3903 bp in length
28151: gap of unknown length
28152: contig of 3562 bp in length
32155: gap of unknown length
35717: contig of 5075 bp in length
40891: gap of unknown length
40991: contig of 3036 bp in length
44027: gap of unknown length
44128: contig of 4772 bp in length
48899: gap of unknown length
48900: contig of 5256 bp in length
54255: gap of unknown length
54256: contig of 4639 bp in length
58994: gap of unknown length
58995: contig of 5296 bp in length
64390: gap of unknown length
70865: contig of 6375 bp in length
70866: gap of unknown length
78667: contig of 7702 bp in length
78668: gap of unknown length
87207: contig of 8440 bp in length
87208: gap of unknown length
97858: contig of 10551 bp in length
97859: gap of unknown length
109520: contig of 11562 bp in length
109521: gap of unknown length
120908: contig of 11288 bp in length
120909: gap of unknown length
141477: contig of 20469 bp in length
141478: gap of unknown length
141578: contig of 21095 bp in length
162672: gap of unknown length
162673: contig of 1202 bp in length
162773: gap of unknown length
163975: contig of 692 bp in length.
164075: Location/Qualifiers
1.164766
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"

FEATURES

source

```
misc_feature      /clone="RP11-245P10"
1. .1124
/notes="assembly_name:Contig14"
misc_feature      1225. .3032
/notes="assembly_name:Contig21"
misc_feature      3133. .4493
/notes="assembly_name:Contig23"
misc_feature      4594. .5860
/notes="assembly_name:Contig24"
misc_feature      5961. .7670
/notes="assembly_name:Contig25"
misc_feature      7771. .9669
/notes="assembly_name:Contig26"
misc_feature      9770. .11715
/notes="assembly_name:Contig27"
misc_feature      11816. .14243
/notes="assembly_name:Contig28"
misc_feature      14344. .16887
/notes="assembly_name:Contig29"
misc_feature      16988. .19247
/notes="assembly_name:Contig30"
misc_feature      19348. .21375
/notes="assembly_name:Contig31"
misc_feature      21476. .25025
/notes="assembly_name:Contig32"
misc_feature      25126. .28051
/notes="assembly_name:Contig33"
misc_feature      28152. .32054
/notes="assembly_name:Contig34"
misc_feature      32155. .35716
/notes="assembly_name:Contig35"
misc_feature      35817. .40891
/notes="assembly_name:Contig36"
misc_feature      40992. .44027
/notes="assembly_name:Contig37"
misc_feature      44128. .48899
/notes="assembly_name:Contig38"
misc_feature      49000. .54255
/notes="assembly_name:Contig39"
misc_feature      54356. .58994
/notes="assembly_name:Contig40"
misc_feature      59095. .64390
/notes="assembly_name:Contig41"
misc_feature      64491. .70865
/notes="assembly_name:Contig42"
misc_feature      70966. .78667
/notes="assembly_name:Contig43"
misc_feature      78768. .87207
/notes="assembly_name:Contig44"
misc_feature      87308. .97858
/notes="assembly_name:Contig45"
misc_feature      97959. .109520
/notes="assembly_name:Contig46"
misc_feature      109621. .120908
/notes="assembly_name:Contig47"
misc_feature      121009. .141477
/notes="assembly_name:Contig48"
misc_feature      141578. .162672
/notes="assembly_name:Contig49"
misc_feature      162773. .163974
/notes="assembly_name:Contig99"
misc_feature      164075. .164766
/notes="assembly_name:Contig7"

Query Match      23.6%; Score 229.4; DB 2; Length 164766;
Best Local Similarity 97.5%; Pred. No. 4.9e-27;
Matches 233; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      191      CTGGCCGAGAGCGGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGTGGCAGATGTTG 250
DB      65348      CTGCCCCCGAGGCGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGTGGCAGATGTTG 65289

ORIGIN

AC023889      174612 bp      DNA      linear      HTG 07-JUL-2000
Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC023889
AC023889.3      GI:8969253
HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
1 (bases 1 to 174612)
The sequence of Homo sapiens clone
Unpublished
Waterston,R.H.
2 (bases 1 to 174612)
Direct Submission
Submitted (18-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8748947.

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0661B12
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 148000; agarose-fp
Insert size: 173412; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.55 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2478: contig of 2478 bp in length
* 2479 2578: gap of unknown length
* 2579 6948: contig of 4370 bp in length
* 6949 7049: gap of unknown length
* 7049 9373: contig of 2325 bp in length
* 9374 9473: gap of unknown length
* 9474 13303: contig of 3830 bp in length
* 13304 13403: gap of unknown length
```

* 13404 18270: contig of 4867 bp in length
* 18271 18370: gap of unknown length
* 18371 25589: contig of 7219 bp in length
* 25590 25689: gap of unknown length
* 25690 34815: contig of 9126 bp in length
* 34816 34915: gap of unknown length
* 34916 45742: contig of 10827 bp in length
* 45743 45842: gap of unknown length
* 45843 62987: contig of 17145 bp in length
* 62988 63087: gap of unknown length
* 63088 87584: contig of 24497 bp in length
* 87585 87684: gap of unknown length
* 87685 114765: contig of 27081 bp in length
* 114766 114865: gap of unknown length
* 114866 142334: contig of 27469 bp in length
* 142335 142434: gap of unknown length
* 142435 174612: contig of 32178 bp in length.

FEATURES

Location/Qualifiers

source

1..174612
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-661B12"

misc_feature

1..2478
/note="assembly_name:Contig4"

misc_feature

2575..6948
/note="assembly_name:Contig5"

misc_feature

7049..9373
/note="assembly_name:Contig6"

misc_feature

9474..13303
/note="assembly_name:Contig7"

misc_feature

13404..18270
/note="assembly_name:Contig8"

misc_feature

18371..25589
/note="assembly_name:Contig9"

misc_feature

25690..34815
/note="assembly_name:Contig10"

misc_feature

34916..45742
/note="assembly_name:Contig11"

misc_feature

45843..62987
/note="assembly_name:Contig12"

misc_feature

63088..87584
/note="assembly_name:Contig13"

misc_feature

87685..114765
/note="assembly_name:Contig14"

misc_feature

114866..142334
/note="assembly_name:Contig15"

misc_feature

142435..174612
/note="assembly_name:Contig16"

ORIGIN

Query Match 23.6%; Score 229.4; DB 2; Length 174612;
Best Local Similarity 97.5%; Pred. No. 4.8e-27;
Matches 233; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 191 CTGGCCGAGAGGGCTCTACTCAGAAATCCGAGGTGAAGGACTACCTGTGGCAGATGTTG 250
DB 19339 CTGCCCCCGAGGGCTCTCTACTCAGAAATCCGAGGTGAAGGACTACCTGTGGCAGATGTTG 19280

QY 251 AGTGCACCCAGTACTCGACAAACAGACATCTCTGACCTGGACCTGGAGTCCGAGAAC 310
DB 19279 AGTGCACCCAGTACTCGACAAACAGACATCTCTGACCTGGACCTGGAGTCCGAGAAC 19220

QY 311 ATGATCATCCGATACACTGCTCAAGTCTGAGCTGGACCTGGGCAATGCACAGGCTC 370
DB 19219 ATGATCATCCGATACACTGCTCAAGTCTGAGCTGGACCTGGGCAATGCACAGGCTC 19160

QY 371 AGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGACCATGGCTCC 429
DB 19159 AGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGACCATGGGTGC 19101

RESULT 15

AX309418
LOCUS AX309418 225 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 2403 from Patent WO0190366.
ACCESSION AX309418
VERSION AX309418.1 GI:17895939
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Leach, M.D. and Shimkets, R.A.
AUTHORS Human polynucleotides and polypeptides encoded thereby
TITLE Patent: WO 0190366-A 2403 29-NOV-2001;
JOURNAL Curagen Corporation (US)
FEATURES
Location/Qualifiers
source
1..225
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 23.0%; Score 223.4; DB 6; Length 225;
Best Local Similarity 99.6%; Pred. No. 1e-25;
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 314 ATCATCACCGAATACAACTGCTCAAGGTCTGTTGGACCTGGGCAATGCACAGAGCCTCAGC 373
DB 1 ATCATCACCGAATACAACTGCTCAAGGTCTGTTGGACCTGGGCAATGCACAGAGCCTCAGC 60

QY 374 CAGGAGAAGTCTGCCCTCAGACACAGTTTCAGGACTACCTAGAGACCATGGCTCCAGAG 433
DB 61 CAGGAGAAGTCTGCCCTCAGACACAGTTTCAGGACTACCTAGAGACCATGGCTCCAGAG 120

QY 434 CTCCTGGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCATCGGTGTGACAGCC 493
DB 121 CTCCTGGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCATCGGTGTGACAGCC 180

QY 494 TTCATCATGCTAGCCCCGAGTACCCGGTGAAGAGGAGGGTGA 538
DB 181 TTCATCATGCTAGCCCCGAGTACCCGGTGAAGAGGAGGGTGA 225

Search completed: September 19, 2004, 19:26:51

Job time : 3971.31 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 12:53:28 / Search time 400.361 Seconds
(without alignments)
10303.209 Million cell updates/sec

Title: US-10-077-130-4_COPY_23150_24120
Perfect score: 971
Sequence: 1 cagcggcgggcgctggcg.....ggcggctaaaaaagtctaga 971

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 971 | 100.0 | 8106 | 7 ABX11641 | Abx11641 Human ser |
| 2 | 971 | 100.0 | 24120 | 7 ABX11642 | Abx11642 Human ser |
| 3 | 946.4 | 97.5 | 4175 | 5 AAC62285 | Aac62285 cDNA enco |
| 4 | 926.8 | 95.4 | 5207 | 6 AAL43909 | Aal43909 Human kin |
| 5 | 926.8 | 95.4 | 5207 | 6 AAL43908 | Aal43908 Human kin |
| 6 | 911.4 | 93.9 | 3225 | 6 AAD26467 | Aad26467 Human kin |
| 7 | 911.4 | 93.9 | 5454 | 6 AAD38865 | Aad38865 Human kin |
| 8 | 905.4 | 93.3 | 4936 | 4 AAH46904 | Aah46904 cDNA enco |
| 9 | 905.4 | 93.2 | 5007 | 5 AAC62286 | Aac62286 cDNA enco |
| 10 | 905.4 | 93.2 | 7928 | 5 AAC62287 | Aac62287 cDNA enco |
| 11 | 223.4 | 23.0 | 225 | 6 ABN76255 | Abn76255 Human kin |
| 12 | 206 | 21.2 | 7710 | 4 AAF44862 | Aaf44862 Novel pro |
| 13 | 204.4 | 21.1 | 7789 | 6 AAD30565 | Aad30565 Human kin |
| 14 | 204.4 | 21.1 | 9807 | 9 AAD79958 | Aad79958 Human kin |
| 15 | 204.4 | 21.1 | 10662 | 9 ADC99127 | Adc99127 Human KPP |
| 16 | 202.8 | 20.9 | 9698 | 9 ADE47675 | Ade47675 Human NOV |
| 17 | 202.8 | 20.9 | 10122 | 9 ADE47673 | Ade47673 Human NOV |
| 18 | 135.6 | 14.0 | 9930 | 9 ADE47671 | Ade47671 Human NOV |
| 19 | 126.8 | 13.1 | 182 | 3 AAC25654 | Aac25654 Human sec |
| 20 | 113.8 | 11.7 | 2079 | 4 AAK94258 | Aak94258 Human ful |
| 21 | 113.8 | 11.7 | 2224 | 4 AAH16158 | Aah16158 Human cDN |
| 22 | 113.8 | 11.7 | 2224 | 5 AAH78068 | Aah78068 Nucleotid |
| 23 | 113.8 | 11.7 | 2226 | 9 ADD29770 | Add29770 Human tum |

| | | | | | |
|----|-------|------|------|------------|--------------------|
| 24 | 112.2 | 11.6 | 2105 | 6 ABV72291 | Abv72291 Nucleotid |
| 25 | 112.2 | 11.6 | 2132 | 2 AAX34656 | Aax34656 Human ZIP |
| 26 | 112.2 | 11.6 | 2132 | 7 ACA90235 | Ac90235 Deatch as |
| 27 | 94.4 | 9.7 | 1514 | 9 ADB58463 | Adb58463 Toxicity- |
| 28 | 94.4 | 9.7 | 1514 | 9 ADB53052 | Adb53052 Primary r |
| 29 | 94.2 | 9.7 | 1429 | 2 AAX34657 | Aax34657 Murine Zi |
| 30 | 90.4 | 9.3 | 757 | 4 AAK93262 | Aak93262 Human cDN |
| 31 | 90.4 | 9.3 | 757 | 4 AAK91856 | Aak91856 Human cDN |
| 32 | 89.2 | 9.2 | 1791 | 6 AAD36211 | Aad36211 Human car |
| 33 | 89.2 | 9.2 | 1791 | 6 AAD36212 | Aad36212 Human car |
| 34 | 89.2 | 9.2 | 1791 | 6 AAD36209 | Aad36209 Human car |
| 35 | 89.2 | 9.2 | 1791 | 6 AAD36040 | Aad36040 Human car |
| 36 | 89.2 | 9.2 | 1791 | 6 AAD36210 | Aad36210 Human car |
| 37 | 89.2 | 9.2 | 1836 | 4 AAS06721 | Aas06721 Polynucle |
| 38 | 89.2 | 9.2 | 2046 | 5 AAH78263 | Aah78263 Nucleotid |
| 39 | 89 | 9.2 | 860 | 9 ADE47677 | Ade47677 Human NOV |
| 40 | 87.6 | 9.0 | 1839 | 6 AAD26573 | Aad26573 Human POL |
| 41 | 84.6 | 8.7 | 1788 | 6 AAD26570 | Aad26570 Human POL |
| 42 | 84.6 | 8.7 | 1971 | 6 AAD26571 | Aad26571 Human POL |
| 43 | 84.6 | 8.7 | 2558 | 6 AAD26572 | Aad26572 Human POL |
| 44 | 83.2 | 8.6 | 1032 | 4 AAS06711 | Aas06711 Polynucle |
| 45 | 83.2 | 8.6 | 1216 | 6 ABZ11323 | Abz11323 Human poi |

ALIGNMENTS

RESULT 1
ABX11641
ID ABX11641 standard; cDNA; 8106 BP.
XX
AC ABX11641;
XX
DT 09-MAY-2003 (first entry)
XX
DE Human serine/threonine or protein kinase 59079, cDNA.
XX
KW Human; ss; gene; serine/threonine kinase; protein kinase; 59079;
KW cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
KW haemolytic anaemia; cellular proliferative disorder; cancer;
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..71
FT CDS /*tag= a
FT /*tag= b
FT /*product= "Kinase 59079"
FT /*note= "This CDS is specifically claimed in claim 2"
FT 3'UTR 7965..8106
FT /*tag= c
FT US2002168742-A1.
FT 14-NOV-2002.
FT 15-FEB-2002; 2002US-00077130.
FT 15-FEB-2001; 2001US-0269201P.
FT (WILL-) MILLENNIUM PHARM INC.
FT Kapeller-Libermann R, Acton SL;
FT WPI; 2003-298729/29.
FT P-PSDB; ABG76186.
XX

PT Novel isolated human protein kinase, designated 59079 or 12599
 PT polypeptide, useful as diagnostic and therapeutic agents for preventing
 PT cardiovascular diseases, proliferative disorders, and protein kinase
 PS disorders.

XX Claim 2; Page 39-48; 119pp; English.

CC The invention relates to an isolated human serine/threonine or protein
 CC kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule
 CC comprising at least 85% identity to the nucleic acids appearing as
 CC ABX11641 and ABX11642 or their complement, a naturally occurring variant
 CC of the kinases or their fragments. Also included are a non-human host
 CC cell containing the nucleic acids, an antibody specific for the proteins,
 CC identifying a compound which binds to the kinase (by contacting the
 CC kinase or a cell expressing the kinase with a test compound and
 CC determining whether the kinase binds to the test compound) and modulating
 CC the activity of kinase using the identified compound. The kinases and
 CC their encoding nucleic acids are useful as diagnostic and therapeutic
 CC agents for preventing a disease or condition associated with an aberrant
 CC or unwanted 59079 or 12599 activity in a subject, including
 CC cardiovascular diseases such as heart failure, and myocardial infarction;
 CC disorders involving blood vessels such as atherosclerosis, and Kaposi's
 CC sarcoma; blood platelets disorder such as thrombocytopenia, leukaemia,
 CC Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders
 CC such as cancer; and protein kinase disorders such as autoimmune
 CC disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,
 CC rheumatoid arthritis, and multiple sclerosis (many examples of diseases
 CC and disorders are included in the specification). The kinases, their
 CC encoding nucleic acids and antibodies are useful in screening assays,
 CC detection assays (e.g. forensic biology), and predictive medicine (e.g.
 CC diagnostic assays, prognostic assays, and monitoring clinical trials and
 CC pharmacogenomics). The kinases and their encoding nucleic acids are
 CC useful as query sequences to perform a search against public databases to
 CC identify other family members or related sequences. The present sequence
 CC encodes the kinase 59079

XX Sequence 8106 BP; 1503 A; 2684 C; 2530 G; 1289 T; 0 U; 0 Other;

Query Match 100.0%; Score 971; DB 7; Length 8106;
 Best Local Similarity 100.0%; Pred. No. 1.2e-171;
 Matches 971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCGGGGGCGCTGGCCGCGCAGATCATCCCTACCCACCACCAAGGACAGACAGT 60
 DB 7136 CAGCGGGGGCGCTGGCCGCGCAGATCATCCCTACCCACCACCAAGGACAGACAGT 7195
 QY 61 GCTCGCGGAATACAGAGCCCTCAAGGGCCCTGGCCACCCGACCTGGCCAGTGCACGC 120
 DB 7196 GCTCGCGGAATACAGAGCCCTCAAGGGCCCTGGCCACCCGACCTGGCCAGTGCACGC 7255
 QY 121 AGCTACTCTAGCCCGGACCTGGTCTCATCTTGGAGCTGTCTGGGCGCCAGCT 180
 DB 7256 AGCTACTCTAGCCCGGACCTGGTCTCATCTTGGAGCTGTCTGGGCGCCAGCT 7315
 QY 181 GCTCCCTCTGCTGGCCGAGAGGGCTCTACTCAGATCCGAGTGAAGACTACCTGTG 240
 DB 7316 GCTCCCTCTGCTGGCCGAGAGGGCTCTACTCAGATCCGAGTGAAGACTACCTGTG 7375
 QY 241 GCAGATGTGATGTCACCCAGTACCTGCAACACAGACATCTGTGACCTGACCTGAG 300
 DB 7376 GCAGATGTGATGTCACCCAGTACCTGCAACACAGACATCTGTGACCTGACCTGAG 7435
 QY 301 GTCCGAGAAATGATCATACCGAATCAACCTGCTCAAGTCTGTGACCTGGGCAATGC 360
 DB 7436 GTCCGAGAAATGATCATACCGAATCAACCTGCTCAAGTCTGTGACCTGGGCAATGC 7495
 QY 361 ACAGAGCTTACGACAGAGAGTGTGCTGCTCAGCAAGATTCAAGACTACCTAGAGAC 420
 DB 7496 ACAGAGCTTACGACAGAGAGTGTGCTGCTCAGCAAGATTCAAGACTACCTAGAGAC 7555
 QY 421 CATGGCTCCAGAGCTCTCGAGGGGCCAGGGGGCTGTTCACACAGACAGATCTGGGCCAT 480
 DB 7556 CATGGCTCCAGAGCTCTCGAGGGGCCAGGGGGCTGTTCACACAGACAGATCTGGGCCAT 7615

QY 481 CGGTGTGACAGCCCTTCATCATGCTGAGCGCGGAGTACCCGCTGAGCAGGAGGTGCACG 540
 DB 7616 CGGTGTGACAGCCCTTCATCATGCTGAGCGCGGAGTACCCGCTGAGCAGGAGGTGCACG 7675
 QY 541 CGACCTGCAGAGAGACTGCGCAAGGGGCTGGTCCGGCTGAGCCGCTGCTACTACGCGGGCT 600
 DB 7676 CGACCTGCAGAGAGACTGCGCAAGGGGCTGGTCCGGCTGAGCCGCTGCTACTACGCGGGCT 7735
 QY 601 GTCCGGGGGGCGGCTGGGCTTCTTCGCGAGCACTCTGTGCGCCGACGCTGGGGCCGGCC 660
 DB 7736 GTCCGGGGGGCGGCTGGGCTTCTTCGCGAGCACTCTGTGCGCCGACGCTGGGGCCGGCC 7795
 QY 661 CTGCGCGTCCAGCTGCTTCGCTGAGTGGCTAACAGAGAGGGCCCGGCTGTTCGCG 720
 DB 7796 CTGCGCGTCCAGCTGCTTCGCTGAGTGGCTAACAGAGAGGGCCCGGCTGTTCGCG 7855
 QY 721 GCCCGCGCGCTGAGCTTCCCTACCGCGGCTGGGCTTCTTCGCGCAATCCGCGAA 780
 DB 7856 GCCCGCGCGCTGAGCTTCCCTACCGCGGCTGGGCTTCTTCGCGCAATCCGCGAA 7915
 QY 781 GAGAGCGCGCTGTGTACAGAGGACCAACTGGCCAGGTGGCTGAGGCTGCCCGC 840
 DB 7916 GAGAGCGCGCTGTGTACAGAGGACCAACTGGCCAGGTGGCTGAGGCTGCCCGC 7975
 QY 841 GCCACACCTTGGTCTCCCTCGGCTGGGCTGCTGCGACGCGCCCAATAAAGCGCCGCTAA 900
 DB 7976 GCCACACCTTGGTCTCCCTCGGCTGGGCTGCTGCGACGCGCCCAATAAAGCGCCGCTAA 8035
 QY 901 CGGGCGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 960
 DB 8036 CGGGCGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8095
 QY 961 AAAAGTCTAGA 971
 DB 8096 AAAAGTCTAGA 8106

RESULT 2
 ABX11642
 ID ABX11642 standard; cDNA; 24120 BP.

XX AC ABX11642;
 XX DT 09-MAY-2003 (first entry)
 XX DE Human serine/threonine or protein kinase 12599, cDNA.
 XX KW Human; ss; gene; serine/threonine kinase; protein kinase; 12599;
 KW cardiovascular disease; heart failure; myocardial infarction;
 KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
 KW blood platelet disorder; thrombocytopenia; leukaemia; Hodgkin's disease;
 KW haemolytic anaemia; cellular proliferative disorder; cancer;
 KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
 KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
 KW multiple sclerosis.
 XX KW Homo sapiens.
 XX OS
 XX FH Key
 XX FT 5'UTR
 XX FT CDS
 XX FT FT
 XX FT FT
 XX FT FT
 XX FT 3'UTR
 XX PN US2002168742-A1.
 XX PD 14-NOV-2002.

Location/Qualifiers
 1..71
 /*tag= a
 72..23978
 /*tag= b
 /product= "Kinase 12599"
 /note= "This CDS is specifically claimed in claim 2"
 23979..24120
 /*tag= c

PF 15-FEB-2002; 2002US-00077130.
 XX
 PR 15-FEB-2001; 2001US-0269201P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Kapeller-Libermann R, Acton SL;
 XX
 DR WPI; 2003-298729/29.
 DR P-PSDB; ABG76187.
 XX
 PT Novel isolated human protein kinase, designated 59079 or 12599
 PT polypeptide, useful as diagnostic and therapeutic agents for preventing
 PT cardiovascular diseases, proliferative disorders, and protein kinase
 PT disorders.
 XX
 PS Claim 2; Page 58-84; 119pp; English.
 CC The invention relates to an isolated human serine/threonine or protein
 CC kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule
 CC comprising at least 85% identity to the nucleic acids appearing as
 CC ABX11641 and ABX11642 or their complement, a naturally occurring variant
 CC of the kinases or their fragments. Also included are a non-human host
 CC cell containing the nucleic acids, an antibody specific for the proteins,
 CC identifying a compound which binds to the kinase (by contacting the
 CC kinase or a cell expressing the kinase with a test compound and
 CC determining whether the kinase binds to the test compound) and modulating
 CC the activity of kinase using the identified compound. The kinases and
 CC their encoding nucleic acids are useful as diagnostic and therapeutic
 CC agents for preventing a disease or condition associated with an aberrant
 CC or unwanted 59079 or 12599 activity in a subject, including
 CC cardiovascular diseases such as heart failure, and myocardial infarction;
 CC disorders involving blood vessels such as atherosclerosis, and Kaposi's
 CC sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia,
 CC Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders
 CC such as cancer; and protein kinase disorders such as autoimmune
 CC rheumatoid arthritis, and multiple sclerosis (many examples of diseases
 CC and disorders are included in the specification). The kinases, their
 CC encoding nucleic acids and antibodies are useful in screening assays,
 CC detection assays (e.g. forensic biology), and predictive medicine (e.g.
 CC diagnostic assays), prognostic assays, and monitoring clinical trials and
 CC pharmacogenomics). The kinases and their encoding nucleic acids are
 CC useful as query sequences to perform a search against public databases to
 CC identify other family members or related sequences. The present sequence
 CC encodes the kinase 12599
 XX
 SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;

QY 301 GTCCGAGAACATGATCATCATCCGATATCAACCTGCTCAAGTCTGCTGGACCTGGCAATGC 360
 DB 23450 GTCCGAGAACATGATCATCACCGAATACAACTGCTCAAGTCTGCTGGACCTGGCAATGC 23509
 QY 361 ACAGAGCCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACTAGAGAC 420
 DB 23510 ACAGAGCCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACTAGAGAC 23569
 QY 421 CATGGCTCCAGAGCTCTCTGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480
 DB 23570 CATGGCTCCAGAGCTCTCTGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 23629
 QY 481 CGGTGTGACAGCTTTCATCATCTGAGCCGCGAGTACCCGGTGTGAGCAGGAGGTGCACG 540
 DB 23630 CGGTGTGACAGCTTTCATCATCTGAGCCGCGAGTACCCGGTGTGAGCAGGAGGTGCACG 23689
 QY 541 CGACCTGCAGAGGAGTCTCCGCAAGGGGCTGTGCTCCGGTGTGAGCCGCTGTCTACCGGGCT 600
 DB 23690 CGACCTGCAGAGGAGTCTCCGCAAGGGGCTGTGCTCCGGTGTGAGCCGCTGTCTACCGGGCT 23749
 QY 601 GTCCGGGGGGCGGCTGCTTCTCCGCGAGCACTCTGTGCGCCCGAGCCCTGGGGCGGCGC 660
 DB 23750 GTCCGGGGGGCGGCTGCTTCTCCGCGAGCACTCTGTGCGCCCGAGCCCTGGGGCGGCGC 23809
 QY 661 CTGCGCGTCCAGCTGCTGCTGAGTCCCGTGGCTTAACAGAGGAGGGCCCGGCTGTTCGCG 720
 DB 23810 CTGCGCGTCCAGCTGCTGCTGAGTCCCGTGGCTTAACAGAGGAGGGCCCGGCTGTTCGCG 23869
 QY 721 GCCCGCGCGCTGACCTTCCCTACCGCGCGGCTGTGCGCTTCTGCTGCGCAATCGCGAGAA 780
 DB 23870 GCCCGCGCGCTGACCTTCCCTACCGCGCGGCTGTGCGCTTCTGCTGCGCAATCGCGAGAA 23929
 QY 781 GAGACCGCGCTGCTCTCAAGAGGCAACAACCTGGCCCGAGGTGCGCTGAGGGTTCGCCCCG 840
 DB 23930 GAGACCGCGCTGCTCTCAAGAGGCAACAACCTGGCCCGAGGTGCGCTGAGGGTTCGCCCCG 23989
 QY 841 GCCACACCTTGTCTCCCGTGGGGTGTGCTGAGAGCGCGCAATAAAGCCGCGCAGC 900
 DB 23990 GCCACACCTTGTCTCCCGTGGGGTGTGCTGAGAGCGCGCAATAAAGCCGCGCAGC 24049
 QY 901 CGGGCGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 960
 DB 24050 CGGGCGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 24109
 QY 961 AAAAGTCTAGA 971
 DB 24110 AAAAGTCTAGA 24120

RESULT 3
 AAC62285
 ID AAC62285 standard; cDNA; 4175 BP.
 XX
 AC AAC62285;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE cDNA encoding a human signal transduction polypeptide.
 XX
 KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
 KW congestive heart failure; dilated congestive cardiomyopathy;
 KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
 KW mitral valve disease; aortic valve disease; tricuspid valve disease;
 KW myocardial infarction; cardiac arrhythmia; atherosclerosis;
 KW atherosclerosis; cardiac tumour; microbial infection; ss.
 XX
 OS Homo sapiens.
 XX
 PH Location/Qualifiers
 FT 1..4056
 FT /*tag= a
 FT /product= "signal transduction polypeptide H19G5"
 XX

PN WO200063381-A1.
 XX 26-OCT-2000.
 XX 11-APR-2000; 2000WO-US009488.
 XX 16-APR-1999; 99US-0129553P.
 XX (SCIO-) SCTOS INC.
 PA Zeng W, Stanton L, Kong H;
 PI WPI; 2001-007013/01.
 DR P-PSDB; AAB30567.
 XX Novel H19G5 polypeptides capable of regulating signal transduction and
 PT exhibiting kinase activity useful for identifying antibodies to treat
 PT cardiac diseases, and additional mediators of signal transduction.
 XX
 PS Claim 4; Page 57-59; 81pp; English.
 CC The present sequence encodes a human protein with putative function in
 CC signal transduction. The polypeptide is designated H19G5. The protein is
 CC capable of regulating signal transduction and exhibits kinase activity.
 CC The H19G5 transcript is expressed in the heart. H19G5 polypeptides and
 CC polynucleotides are useful for preventing or treating a cardiac disease,
 CC such as congestive heart failure, dilated congestive cardiomyopathy,
 CC hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve
 CC disease, aortic valve disease or tricuspid valve disease, angina
 CC pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial
 CC or renovascular hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours in humans. The polypeptide is also useful for detecting
 CC the expression of a protein capable of regulating signal transduction or
 CC the expression of a protein capable of acting as a donor or acceptor
 CC molecule of a phosphate group. The monoclonal antibodies can be used as
 CC probes for detecting discrete antigens expressed by tissue or cell
 CC samples, and therefore used in humans for localization and monitoring of
 CC microbial infection
 XX
 SQ Sequence 4175 BP; 778 A; 1459 C; 1283 G; 655 T; 0 U; 0 Other;
 Query Match 97.5%; Score 946.4; DB 5; Length 4175;
 Best Local Similarity 99.9%; Pred. No. 4.1e-167;
 Matches 947; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CAGCGGGCGGCGTGGCCGCAAGATCATCCCTACACCCCAAGCAAGACAGCAGT 60
 DB 3228 CAGCGGGCGGCGTGGCCGCAAGATCATCCCTACACCCCAAGCAAGACAGCAGT 3287
 QY 61 GCTGGCGCAATACAGGCGCTCAAGGCGCTGGCCGCAAGATCATCCCTACACCCCAAGCAAGACAGCAGT 120
 DB 3288 GCTGGCGCAATACAGGCGCTCAAGGCGCTGGCCGCAAGATCATCCCTACACCCCAAGCAAGACAGCAGT 3347
 QY 121 AGCCTACTCAGCCCCCGGCAACCTGGTCTCATCTTGGAGCTGTGCTCTGGCGCCGAGCT 180
 DB 3348 AGCCTACTCAGCCCCCGGCAACCTGGTCTCATCTTGGAGCTGTGCTCTGGCGCCGAGCT 3407
 QY 181 GCTCCCTGCTGGCGGAGAGGCTCTCTACTAGAAATCGAGGTGAAGACTACCTGTG 240
 DB 3408 GCTCCCTGCTGGCGGAGAGGCTCTCTACTAGAAATCGAGGTGAAGACTACCTGTG 3467
 QY 241 GCAGATGTGAGTGCACCCAGTACCTGCACCAACAGCAGCATCTCTGACCTGGACCTGAG 300
 DB 3468 GCAGATGTGAGTGCACCCAGTACCTGCACCAACAGCAGCATCTCTGACCTGGACCTGAG 3527
 QY 301 GTCCGAGAAATGATCATCAGCAATACACCTGCTCAAGGTGTGACCTGGGCAATGC 360
 DB 3528 GTCCGAGAAATGATCATCAGCAATACACCTGCTCAAGGTGTGACCTGGGCAATGC 3587
 QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTCAAGGACTACCTAGAGAC 420
 DB 3588 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTCAAGGACTACCTAGAGAC 3647

QY 421 CATGGCTCCAGAGCTCTGGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCCAT 480
 DB 3648 CATGGCTCCAGAGCTCTGGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCCAT 3707
 QY 481 CGGTGTGACAGGCTTTCATCATCTGAGGCGCCAGGTACCGGTGAGCAGCGGTGCACG 540
 DB 3708 CGGTGTGACAGGCTTTCATCATCTGAGGCGCCAGGTACCGGTGAGCAGCGGTGCACG 3767
 QY 541 CGACCTGACAGAGACTGCGCAAGGGGCTGTTCGGCTGAGCCGCTGCTAGCGGGGCT 600
 DB 3768 CGACCTGACAGAGACTGCGCAAGGGGCTGTTCGGCTGAGCCGCTGCTAGCGGGGCT 3827
 QY 601 GTCCGGGGGCGCGCTGGGCTTCTCTCGAGCACTCTGTGCGCCCAAGCCCTGGGGCGGGCC 660
 DB 3828 GTCCGGGGGCGCGCTGGGCTTCTCTCGAGCACTCTGTGCGCCCAAGCCCTGGGGCGGGCC 3887
 QY 661 CTGGCGCTCCAGCTGCTGAGTCCCGTGGCTAACAGAGAGGCGCCGCGCTGTTCGCG 720
 DB 3888 CTGGCGCTCCAGCTGCTGAGTCCCGTGGCTAACAGAGAGGCGCCGCGCTGTTCGCG 3947
 QY 721 GCCCGGCGCGCTGACCTTCCCTACCGCGCGCTGCGCTCTTCGTGCGCAATCGCGAA 780
 DB 3948 GCCCGGCGCGCTGACCTTCCCTACCGCGCGCTGCGCTCTTCGTGCGCAATCGCGAA 4007
 QY 781 GAGAGCGCGCTGCTGTATACAGAGGCAACAACCTGGCCAGGTGCGCTGAGGTGCGCCCG 840
 DB 4008 GAGAGCGCGCTGCTGTATACAGAGGCAACAACCTGGCCAGGTGCGCTGAGGTGCGCCCG 4067
 QY 841 GCCACACCTTGGTCTCCCGCTGGGCTGCGCTGAGAGCGCGCAATAAAGCCCGCAGC 900
 DB 4068 GCCACACCTTGGTCTCCCGCTGGGCTGCGCTGAGAGCGCGCAATAAAGCCCGCAGC 4127
 QY 901 CGGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 948
 DB 4128 CGGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 4175
 RESULT 4
 AAL43909
 ID AAL43909 standard; DNA; 5207 BP.
 XX
 AC AAL43909;
 XX
 DT 19-SEP-2002 (first entry)
 XX
 DE Human kinase protein coding sequence 2.
 XX
 KW Human; gene; ds; Gene therapy; chromosome 1; kinase protein;
 KW myosin light chain kinase subfamily; kinase protein-mediated disease;
 KW transgenic animal.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 109..5106
 FT /product= "Human kinase protein"
 FT /tag= a
 FT /replace(311, C)
 FT /tag= c
 FT /standard_name= "Single nucleotide polymorphism"
 FT /replace(311, T)
 FT /tag= b
 FT /standard_name= "Single nucleotide polymorphism"
 FT /replace(1741, T)
 FT /tag= d
 FT /standard_name= "Single nucleotide polymorphism"
 FT /replace(2714, C)
 FT /tag= e
 FT /standard_name= "Single nucleotide polymorphism"
 FT /replace(2745, T)
 FT /tag= f
 FT /standard_name= "Single nucleotide polymorphism"
 FT /replace(2859, A)
 FT /tag= g
 FT /standard_name= "Single nucleotide polymorphism"

```
FT      /tag= g
FT      /standard_name= "Single nucleotide polymorphism"
FT      variation :
FT      replace(3420, C)
FT      /tag= h
FT      /standard_name= "Single nucleotide polymorphism"
XX      WO200240683-A2.
XX      23-MAY-2002:
XX      22-OCT-2001; 2001WO-US032616.
XX      14-NOV-2000; 2000US-00711134.
XX      17-MAY-2001; 2001US-00858664.
XX      (PEKE ) PE CORP NY.
XX      Wei M, Ketchum K, Di Francesco V, Beasley EM;
XX      WPI; 2002-500223/53.
XX      P-PSDB; AAO15372.
XX      New kinase proteins related to myosin light chain kinase subfamily and
XX      encoding polynucleotide, useful for diagnosing, treating disease or
XX      condition mediated by the kinase protein and for identifying modulators.
XX      Claim 23; Fig 3; 96pp; English.
XX      The invention comprises the amino acid and coding sequences (located on
XX      chromosome 1) of a human kinase protein that is related to the myosin
XX      light chain kinase subfamily. The human kinase DNA and protein sequences
XX      of the invention are useful for identifying agents that modulate the
XX      activity of the human kinase protein. Kinase-modulating agents are useful
XX      for treating a disease or condition mediated by a human kinase protein.
XX      The human kinase DNA sequences can be used to produce transgenic animals
XX      which are useful for studying the function of kinase proteins and
XX      identifying/evaluating modulators of kinase protein activity. The present
XX      DNA sequence encodes the human kinase protein of the invention.
XX      Sequence 5207 BP; 988 A; 1758 C; 1612 G; 849 T; 0 U; 0 Other;
XX      Query Match      95.4%; Score 926.8; DB 6; Length 5207;
XX      Best Local Similarity 99.8%; Pred. No. 1.9e-163;
XX      Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX      QY      1 CAGCGGGGGGGCGTGGCGCCCAAGATCATCCCTACACCCCAAGGACAGACAGCAGT 60
XX      DB      4278 CAGCGGGGGGGCGTGGCGCCCAAGATCATCCCTACACCCCAAGGACAGACAGCAGT 4337
XX      QY      61 GCTGCGGAATACGAGGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCGACAGTGCACGC 120
XX      DB      4338 GCTGCGGAATACGAGGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCGACAGTGCACGC 4397
XX      QY      121 AGCTTACCTACGCCCCCGGACCTGGTGTCTATCTTGGAGCTGTGCTTGGCCCGGAGCT 180
XX      DB      4398 AGCTTACCTACGCCCCCGGACCTGGTGTCTATCTTGGAGCTGTGCTTGGCCCGGAGCT 4457
XX      QY      181 GCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGAACTCCGAGGTGAAGGACTACCTCTGT 240
XX      DB      4458 GCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGAACTCCGAGGTGAAGGACTACCTCTGT 4517
XX      QY      241 GCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGACATCTCTGCACTGGACCTGAG 300
XX      DB      4518 GCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGACATCTCTGCACTGGACCTGAG 4577
XX      QY      301 GTCGAGAACATGATCATCCGAATACACCTGCTCAAGTCTGAGTCTGGACCTGGCAATGC 360
XX      DB      4578 GTCGAGAACATGATCATCCGAATACACCTGCTCAAGTCTGAGTCTGGACCTGGCAATGC 4637
XX      QY      361 ACAGAGCCTCAGCAGGAGAGAGTGTGCTCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420
XX      DB      4638 ACAGAGCCTCAGCAGGAGAGAGTGTGCTCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 4697
```

```
QY      421 CATGGCTCCAGAGCTCTCTGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCCAT 480
DB      4698 CATGGCTCCAGAGCTCTCTGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCCAT 4757
QY      481 CGGTGTGACAGACCTTCATCATGTCAGCCGCCAGTACCCGGTGAAGAGAGAGGTGCACG 540
DB      4758 CGGTGTGACAGACCTTCATCATGTCAGCCGCCAGTACCCGGTGAAGAGAGAGGTGCACG 4817
QY      541 CGACCTTGACAGAGAGACTCGCCAAAGGGCTGTGTCGGGTGAGCCGCTGTACCGGGGT 600
DB      4818 CGACCTTGACAGAGAGACTCGCCAAAGGGCTGTGTCGGGTGAGCCGCTGTACCGGGGT 4877
QY      601 GTCCGGGGGGGGCGTGGCCCTTCTGCGCAGACACTCTGTGCGCCGACCCCTGGGGCGGGCC 660
DB      4878 GTCCGGGGGGGGCGTGGCCCTTCTGCGCAGACACTCTGTGCGCCGACCCCTGGGGCGGGCC 4937
QY      661 CTGCGGCTCGAGCTGCTGTCAGTGCCTGTGGTGTAAACAGAGAGAGGGCCCGCTGTTCGCG 720
DB      4938 CTGCGGCTCGAGCTGCTGTCAGTGCCTGTGGTGTAAACAGAGAGAGGGCCCGCTGTTCGCG 4997
QY      721 GCCCGCGCGGTGACCTTCCCTACCGCGGGCTGCGGCTCTTCGTGCGCAATCGCGAGAA 780
DB      4998 GCCCGCGCGGTGACCTTCCCTACCGCGGGCTGCGGCTCTTCGTGCGCAATCGCGAGAA 5057
QY      781 GAGACGCGCGTCTGCTGTACAAGAGGCAACAACCTGGCCGAGGTGCGCTGAGGGTTCGCCCG 840
DB      5058 GAGACGCGCGTCTGCTGTACAAGAGGCAACAACCTGGCCGAGGTGCGCTGAGGGTTCGCCCG 5117
QY      841 GCCACACCTTGTCTCCCGGTGGGGTCTGCTGACAGACGCGCAATAAAGCCGCGCAGC 900
DB      5118 GCCACACCTTGTCTCCCGGTGGGGTCTGCTGACAGACGCGCAATAAAGCCGCGCAGC 5177
QY      901 CGGGCGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 930
DB      5178 CGGGCGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5207
XX      RESULT 5
XX      AAL43908
XX      ID AAL43908 standard; DNA; 5207 BP.
XX      AC AAL43908;
XX      DT 19-SEP-2002 (first entry)
XX      DE Human kinase protein coding sequence 1.
XX      KW Human; gene; ds; gene therapy; chromosome 1; kinase protein;
XX      KW myosin light chain kinase subfamily; kinase protein-mediated disease;
XX      KW transgenic animal.
XX      OS Homo sapiens.
XX      FH Key Location/Qualifiers
XX      CDS 109..5106
XX      FT /*tag= a
XX      FT /product= "Human kinase protein"
XX      WO200240683-A2.
XX      23-MAY-2002.
XX      22-OCT-2001; 2001WO-US032616.
XX      14-NOV-2000; 2000US-00711134.
XX      17-MAY-2001; 2001US-00858664.
XX      (PEKE ) PE CORP NY.
XX      Wei M, Ketchum K, Di Francesco V, Beasley EM;
XX      WPI; 2002-500223/53.
XX      P-PSDB; AAO15372.
XX      DR
```

XX New kinase proteins related to myosin light chain kinase subfamily and
 PT encoding polynucleotide, useful for diagnosing, treating disease or
 PT condition mediated by the kinase protein and for identifying modulators.

XX Claim 23; Fig 1; 96pp; English.

XX The invention comprises the amino acid and coding sequences (located on
 CC chromosome 1) of a human kinase protein that is related to the myosin
 CC light chain kinase subfamily. The human kinase DNA and protein sequences
 CC of the invention are useful for identifying agents that modulate the
 CC activity of the human kinase protein. Kinase-modulating agents are useful
 CC for treating a disease or condition mediated by a human kinase protein.
 CC The human kinase DNA sequences can be used to produce transgenic animals
 CC which are useful for studying the function of kinase proteins and
 CC identifying/evaluating modulators of kinase protein activity. The present
 CC DNA sequence encodes the human kinase protein of the invention

XX Sequence 5207 BP; 988 A; 1758 C; 1612 G; 849 T; 0 U; 0 Other;

XX Query Match 95.4%; Score 926.8; DB 6; Length 5207;

XX Best Local Similarity 99.8%; Pred. No. 1.9e-163;

XX Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGCGCGCGGCGCTGGCGGCGGAGATCATCCCTACCCACCCCAAGGACAGACAGT 60

DB 4278 CAGCGCGGCGGCGCTGGCGGCGGAGATCATCCCTACCCACCCCAAGGACAGACAGT 4337

QY 61 GTGCGCGGAATACAGAGGCGCTCAAGGGGCTGCGGCACCCGACCTGCGCCAGCTGCACGC 120

DB 4338 GTGCGCGGAATACAGAGGCGCTCAAGGGGCTGCGGCACCCGACCTGCGCCAGCTGCACGC 4397

QY 121 AGCTTACCTAGCCCCCGGACCTGGTGCTCATCTTGAGAGTGCTGTGGGCCGAGCT 180

DB 4398 AGCTTACCTAGCCCCCGGACCTGGTGCTCATCTTGAGAGTGCTGTGGGCCGAGCT 4457

QY 181 GTCCCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGAAATCCGAGGTGAAGGACTACCTGTG 240

DB 4458 GTCCCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGAAATCTGAGTGAAGGACTACCTGTG 4517

QY 241 CGAGATGTGAGTGCCACCCAGTACCTGCAACACAGACATCTGCACTGGACCTGAG 300

DB 4518 GCAGATGTGAGTGCCACCCAGTACCTGCAACACAGACATCTGCACTGGACCTGAG 4577

QY 301 GTCCGAGAACATGATCATCCGGAATACACCTGCTCAAGGTCTGAGACCTGGGCAATGC 360

DB 4578 GTCCGAGAACATGATCATCCGGAATACACCTGCTCAAGGTCTGAGACCTGGGCAATGC 4637

QY 361 ACAGAGCTTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420

DB 4638 ACAGAGCTTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 4697

QY 421 CATGGCTCCAGACTCTCGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480

DB 4698 CATGGCTCCAGACTCTCGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 4757

QY 481 CGGTGTGACAGCTTCATCATGCTGAGCGCGGAGTACCCGCTGAGCAGCGAGGTGCACG 540

DB 4758 CGGTGTGACAGCTTCATCATGCTGAGCGCGGAGTACCCGCTGAGCAGCGAGGTGCACG 4817

QY 541 CGACTCGAGAGAGACTCCGCAAGGGGCTGGTCCGGCTGAGCGGCTGCTACGGGGGCT 600

DB 4818 CGACTCGAGAGAGACTCCGCAAGGGGCTGGTCCGGCTGAGCGGCTGCTACGGGGGCT 4877

QY 601 GTCCGGGGGCGCGCTGCTTCCTCGCAGACATCTGTGCGGCCAGCCCTGGGGCGGCC 660

DB 4878 GTCCGGGGGCGCGCTGCTTCCTCGCAGACATCTGTGCGGCCAGCCCTGGGGCGGCC 4937

QY 661 GTGGCGCTCAGCTGCTGAGTGGCGGTGCTGAACAGAGAGGGCCCGGCTGTCGG 720

DB 4938 GTGGCGCTCAGCTGCTGAGTGGCGGTGCTGAACAGAGAGGGCCCGGCTGTCGG 4997

QY 721 GCCCGCGCGCTGACCTTCCTACCGCGGCTGCGGCTCTTCGTGCGCAATCGCGAGAA 780

DB 4998 GCCCGCGCGCTGACCTTCCTACCGCGGCTGCGGCTTCCTGTCGCAATCGCGAGAA 5057

QY 781 GAGACGCGCGCTGCTGTACAGAGGACACAACTGCGCCAGGTGCGCTGAGGTGCGCCCG 840

DB 5058 GAGACGCGCGCTGCTGTACAGAGGACACAACTGCGCCAGGTGCGCTGAGGTGCGCCCG 5117

QY 841 GCCACACCTTGGTCTCTCCCGCTGCGGGTCTGCTGAGACGCGCCCAATAAAGCCCGCAGC 900

DB 5118 GCCACACCTTGGTCTCTCCCGCTGCGGGTCTGCTGAGACGCGCCCAATAAAGCCCGCAGC 5177

QY 901 CGGCGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 930

DB 5178 CGGCGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 5207

RESULT 6

AD26467

ID AAD26467 standard; cDNA; 3225 BP.

XX AC AAD26467;

XX AC AAD26467;

DT 26-MAR-2002 (first entry)

XX Human kinase PKIN-20 cDNA.

XX Human; kinase; PKIN-20; cancer; leukaemia; adenocarcinoma; osteoporosis;

KW Immune disorder; atherosclerosis; Crohn's disease; Hodgkin's disease;

KW Acquired Immune Deficiency Syndrome; AIDS; Addison's disease; anaemia;

KW allergy; asthma; adult respiratory distress syndrome; multiple sclerosis;

KW autoimmune thyroiditis; bronchitis; diabetes mellitus; osteoarthritis;

KW Good pasture's syndrome; Graves' disease; pancreatitis; psoriasis;

KW rheumatoid arthritis; ulcerative colitis; cirrhosis; Cushing's syndrome;

KW hepatitis; hypothyroidism; cerebral palsy; cataract; angina pectoris;

KW cardiovascular disease; hypertension; vasculitis; myocarditis; obesity;

KW congestive heart failure; ischaemic heart disease; lung tumour; gout;

KW fatty liver; Niemann-Pick's disease; gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 526..3141

FT /tag= a

FT /product= "Human PKIN-20 protein"

XX WO200196547-A2.

XX 20-DEC-2001.

XX 14-JUN-2001; 2001WO-US019444.

XX 15-JUN-2000; 2000US-0212073P.

XX 23-JUN-2000; 2000US-0213467P.

XX 30-JUN-2000; 2000US-0215651P.

XX 07-JUL-2000; 2000US-0216605P.

XX 13-JUL-2000; 2000US-0218372P.

XX 25-AUG-2000; 2000US-0228056P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Lal P, Bandman O, Borowsky ML, Au-Young J, Lu Y;

PI Gandhi AR, Tribouley CM, Walla NK, Yao MG, Lu DAM, Greenwald SR;

PI Ramkumar J, Griffin JA, Kearney L, Burford N, Nguyen DB, Tang YT;

PI Baughn MR, He A, Thornton M, Rafalia A, Patterson C, Gururajan R;

PI Lo TP, Khan F, Recipon SA, Azimzai Y, Policky JL, Ding L;

PI Grether M, Elliott VS, Thangavelu K, Batra S, Ison CH;

XX WPI; 2002-090207/12.

DR P-PSDB; AAE16274.

XX New polypeptides, useful for diagnosing, treating or preventing disorders

PT of growth and development, cardiovascular and lipid, and diseases such as

PT cancer, comprise human kinase polypeptides.

Claim 5; Page 19; 197pp; English.

The invention relates to human kinase PKIN proteins and their corresponding cDNAs. A composition containing PKIN agonist is useful for treating a disease or condition associated with decreased expression of PKIN and a composition comprising PKIN antagonist is useful for treating a disease or condition associated with overexpression of PKIN. The disorders include cancer (leukaemia, adenocarcinoma, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, Hodgkin's disease); immune disorder (Acquired Immune Deficiency Syndrome (AIDS), asthma, Addison's disease, atherosclerosis, anaemia, allergies, adult respiratory distress syndrome, autoimmune thyroiditis, gout, bronchitis, Crohn's disease, diabetes mellitus, multiple sclerosis, Good pasture's syndrome, Graves' disease, osteoarthritis, osteoporosis, pancreatitis, psoriasis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis, ulcerative colitis, bacterial, parasitic, fungal, viral, protozoal and helminthic infections), growth and development disorders (arteriosclerosis, cirrhosis, hepatitis, Cushing's syndrome, hypothyroidism, cerebral palsy, cataracts); cardio vascular disease (arteriovenous fistula, hypertension, vasculitis, aneurysms, congestive heart failure, angina pectoris, myocarditis, ischaemic heart disease, chronic bronchitis, lung tumours); lipid disorder (fatty liver, Fabry's disease, Niemann-Pick's disease, hypocholesterolaemia, obesity). PKIN DNA is useful for assessing toxicity of a test compound and in gene therapy. The present sequence is human PKIN-20 cDNA

Sequence 3225 BP; 596 A; 1104 C; 1017 G; 508 T; 0 U; 0 Other;

| Query Match | 93.9% | Score 911.4 | DB 6 | Length 3225 |
|-----------------------|----------------|---|----------|-------------|
| Best Local Similarity | 99.9% | Pred. No. 1.3e-160 | | |
| Matches 912 | Conservative 0 | Mismatches 1 | Indels 0 | Gaps 0 |
| QY | 1 | CAGCGGCGGGCGCTGGCCGCCAAGATCATCCCTACACCCCAAGACAAAGACAGCAGT | 60 | |
| Db | 2313 | CAGCGGCGGGCGCTGGCCGCCAAGATCATCCCTACACCCCAAGACAAAGACAGCAGT | 2372 | |
| QY | 61 | GCTGCGGGATACAGAGCCCTCANGGGCTTGGCCACCGCACCTGGCCAGCTGACGC | 120 | |
| Db | 2373 | GCTGCGGGATACAGAGCCCTCANGGGCTTGGCCACCGCACCTGGCCAGCTGACGC | 2432 | |
| QY | 121 | AGCTACTCTCAGCCCCCGGCACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCCCGAGCT | 180 | |
| Db | 2433 | AGCTACTCTCAGCCCCCGGCACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCCCGAGCT | 2492 | |
| QY | 181 | GCTCCCTGCTGTCGCCAGAGGGGCTCTCTACTCAGAATCCGAGGTGAAGGACTACCTGTG | 240 | |
| Db | 2493 | GCTCCCTGCTGTCGCCAGAGGGGCTCTCTACTCAGAATCCGAGGTGAAGGACTACCTGTG | 2552 | |
| QY | 241 | GCAGATGTTGAGTGCCACCCAGTACTGTGCACAAACAGACATCTCTGCACCTGGACCTGAG | 300 | |
| Db | 2553 | GCAGATGTTGAGTGCCACCCAGTACTGTGCACAAACAGACATCTCTGCACCTGGACCTGAG | 2612 | |
| QY | 301 | GTCCGAAACATGATCATCACCGAATACAACTGCTCAAGGTCTGGACCTGGGCAATGC | 360 | |
| Db | 2613 | GTCCGAAACATGATCATCACCGAATACAACTGCTCAAGGTCTGGACCTGGGCAATGC | 2672 | |
| QY | 361 | ACAGAGCTCAGCCAGGAGAAGTGTGCTCCTCAGACAAAGTTCAAGACTACCTTAGAC | 420 | |
| Db | 2673 | ACAGAGCTCAGCCAGGAGAAGTGTGCTCCTCAGACAAAGTTCAAGACTACCTTAGAC | 2732 | |
| QY | 421 | CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGGCTGTTTCCACAGACACATCTGGGCCAT | 480 | |
| Db | 2733 | CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGGCTGTTTCCACAGACACATCTGGGCCAT | 2792 | |
| QY | 481 | CGGTGTGACAGCCTTTCATCTGTCAGGCCGAGTACCCGCTGAGCAGCGGTGACGC | 540 | |
| Db | 2793 | CGGTGTGACAGCCTTTCATCTGTCAGGCCGAGTACCCGCTGAGCAGCGGTGACGC | 2852 | |
| QY | 541 | CGACCTGCAGAGAGGACTGCGCAAGGGGCTGGTCCGGCTGAGCCGCTGCTACGCGGGCT | 600 | |
| Db | 2853 | CGACCTGCAGAGAGGACTGCGCAAGGGGCTGGTCCGGCTGAGCCGCTGCTACGCGGGCT | 2912 | |

| | | | |
|----------|--|---|------|
| Qy | 601 | GTCCGGGGCGCGTGGCTTCTCTGGCAGCACTCTGTGCGCCAGCCCTGGGGCCGGCC | 660 |
| Db | 2913 | GTCCGGGGCGCGTGGCTTCTCTGGCAGCACTCTGTGCGCCAGCCCTGGGGCCGGCC | 2972 |
| Qy | 661 | CTGCGGCTCAGCTGCCTGCAGTGCCTGGCTAAACAGAGAGAGGCGCCGCTGTTGCGG | 720 |
| Db | 2973 | CTGCGGCTCAGCTGCCTGCAGTGCCTGGCTAAACAGAGAGAGGCGCCGCTGTTGCGG | 3032 |
| Qy | 721 | GCCCGGCGCCGCTGACCTTTCCTTACC CGCGCGGCTTCTGTGCGCAATCGCGAGAA | 780 |
| Db | 3033 | GCCCGGCGCCGCTGACCTTTCCTTACC CGCGCGGCTTCTGTGCGCAATCGCGAGAA | 3092 |
| Qy | 781 | GAGACGGCGCTGCTGTACAGAGGACACCTTGGCCCGAGTGCGCTGAGGGTGC GCCCGG | 840 |
| Db | 3093 | GAGACGGCGCTGCTGTACAGAGGACACCTTGGCCCGAGTGCGCTGAGGGTGC GCCCGG | 3152 |
| Qy | 841 | GCCACACCTTGGTCTCCCGCGTGGGGGTGCTGTCAGACGCGCAATAAAAAAGCCGAGC | 900 |
| Db | 3153 | GCCACACCTTGGTCTCCCGCGTGGGGGTGCTGTCAGACGCGCAATAAAAAAGCCGAGC | 3212 |
| Qy | 901 | CGGGCGAGAAAA 913 | |
| Db | 3213 | CGGGCGAGAAAA 3225 | |
| RESULT 7 | | | |
| AAD38865 | | | |
| ID | AAD38865 standard; cDNA; 5454 BP. | | |
| XX | | | |
| AC | AAD38865; | | |
| XX | | | |
| DT | 23-SEP-2002 (first entry) | | |
| XX | | | |
| DE | Human kinase (PKIN)-22 cDNA. | | |
| XX | | | |
| KW | Human; kinase; PKIN; cancer; immune system disorder; atherosclerosis; | | |
| KW | acquired immune deficiency syndrome; AIDS; Addison's disease; allergy; | | |
| KW | asthma; multiple sclerosis; psoriasis; arteriosclerosis; cirrhosis; | | |
| KW | development; hepatitis; cardiovascular; hypertension; drug screening; | | |
| KW | myocardial infarction; Goodpasture's syndrome; lipid disorder; growth; | | |
| KW | fatty liver; Gaucher's disease; Niemann-Pick's disease; anorectic; | | |
| KW | hypercholesterolemia; obesity; gene therapy; cytostatic; anti-HIV; | | |
| KW | neuroprotective; hepatotropic; hypotensive; cardiant; nephrotropic; | | |
| XX | hyperlipidaemia; enzyme; gene; ss. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| PH | Key Location/Qualifiers | | |
| FT | CDS 373..5370 | | |
| FT | /*tag= a | | |
| FT | /product= "Human kinase (PKIN)-22" | | |
| XX | | | |
| FN | WO200233099-A2. | | |
| XX | | | |
| PD | 25-APR-2002. | | |
| XX | | | |
| PF | 20-OCT-2001; 2001WO-US047728. | | |
| XX | | | |
| PR | 20-OCT-2000; 2000US-0242410P. | | |
| PR | 27-OCT-2000; 2000US-0244068P. | | |
| PR | 03-NOV-2000; 2000US-0245708P. | | |
| PR | 09-NOV-2000; 2000US-0247672P. | | |
| PR | 16-NOV-2000; 2000US-0249565P. | | |
| PR | 22-NOV-2000; 2000US-0252730P. | | |
| PR | 01-DEC-2000; 2000US-0250807P. | | |
| XX | | | |
| PA | (INCY-) INCYTE GENOMICS INC. | | |
| XX | | | |
| PI | Gururajan R, Baughn WR, Wallia NK, Elliott VS, Xu Y, Arvizu C; | | |
| PI | Yao MG, Ramkumar J, Ding L, Fang YT, Hafalia AJA, Nguyen DB; | | |
| PI | Fi Gandhi AR, Lu Y, Yue H, Burford N, Bandman O, Tribouley CM, Lal PG; | | |
| PI | Recipon SA, Lu DAM, Borowsky ML, Thornton M, Swarnaker A; | | |
| PI | Thangavelu K, Khan FA, Ison CH; | | |

XX WPI; 2002-454603/48.
 DR P-PSDB; AAE24151.
 XX
 PT New human kinase polypeptide, for diagnosing, preventing and treating
 PT cancer, immune system disorders, growth and development disorders,
 PT cardiovascular disorders and lipid disorders.
 XX
 PS Claim 5; Page 209-210; 210pp; English.
 XX
 CC The invention relates human kinases (PKIN) and their corresponding
 CC nucleic acid sequences. PKIN and its DNA are useful for diagnosing,
 CC treating and preventing cancer, an immune system disorder (e.g., acquired
 CC immune deficiency syndrome (AIDS), Addison's disease, allergy, asthma,
 CC atherosclerosis, multiple sclerosis, psoriasis), disorders affecting
 CC growth and development (e.g., arteriosclerosis, cirrhosis, hepatitis),
 CC cardiovascular disorder (e.g., hypertension, myocardial infarction,
 CC Goodpasture's syndrome), and a lipid disorder (e.g., fatty liver,
 CC Gaucher's disease, Niemann-Pick's disease, hypercholesterolaemia,
 CC hyperlipidaemia, obesity), and for assessing the effects of exogenous
 CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a
 CC condition or a disease associated with the expression of PKIN in a
 CC biological sample. A composition comprising PKIN or an agonist or
 CC antagonist of PKIN is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional PKIN.
 CC PKIN is useful in a number of drug screening techniques and to analyse
 CC the proteome of a tissue or cell type. PKIN DNA is useful for creating
 CC knockin humanised animals or transgenic animals to model human diseases,
 CC and in somatic or germline gene therapy. The present sequence is human
 CC PKIN cDNA
 XX
 SQ Sequence 5454 BP; 1019 A; 1853 C; 1686 G; 896 T; 0 U; 0 Other;

Query Match 93.9%; Score 911.4; DB 6; Length 5454;
 Best Local Similarity 99.9%; Pred. No. 1.4e-160;
 Matches 912; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGCGGCGCTGGCGCGCAAGATCATCCCTACCAACCCCAAGGACAGACAGT 60
 DB 4542 CAGCGGGCGGCGCTGGCGCGCAAGATCATCCCTACCAACCCCAAGGACAGACAGT 4601
 QY 61 GCTGGCGAATACAGAGGCGCTCAAGGGCGCTGGCGCAACCCGACCTGGCCAGCTGCACGC 120
 DB 4602 GCTGGCGAATACAGAGGCGCTCAAGGGCGCTGGCGCAACCCGACCTGGCCAGCTGCACGC 4661
 QY 121 AGCTTACTCAGCCCCCGGACCTGGTGTCTATCTTGGAGCTGTGTCTGGGCCCGAGCT 180
 DB 4662 AGCTTACTCAGCCCCCGGACCTGGTGTCTATCTTGGAGCTGTGTCTGGGCCCGAGCT 4721
 QY 181 GCTCCCTGCTGGCCGAGAGGGCTCTCTACTCAGAACTCCAGGTTGAAGGACTACCTGTG 240
 DB 4722 GCTCCCTGCTGGCCGAGAGGGCTCTCTACTCAGAACTCCAGGTTGAAGGACTACCTGTG 4781
 QY 241 GCAGATGTGATGCGACCCAGTACCTGCAACCCAGACACATCTCCACCTGACCTGAG 300
 DB 4782 GCAGATGTGATGCGACCCAGTACCTGCAACCCAGACACATCTCCACCTGACCTGAG 4841
 QY 301 GTCCGAGAACATGATCATCAACCGAATACAACTGCTCAAGTCTGTGACCTGGGCAATGC 360
 DB 4842 GTCCGAGAACATGATCATCAACCGAATACAACTGCTCAAGTCTGTGACCTGGGCAATGC 4901
 QY 361 ACAGAGCTCAGCAGAGAGAGTGTGCTGCTCAGCAAGTTCAAGGACTACCTAGAGAC 420
 DB 4902 ACAGAGCTCAGCAGAGAGAGTGTGCTGCTCAGCAAGTTCAAGGACTACCTAGAGAC 4961
 QY 421 CATGGCTCCAGAGCTCTCGAGGGGCCAGGGGCTGTTCACACAGACACATCTGGGCCAT 480
 DB 4962 CATGGCTCCAGAGCTCTCGAGGGGCCAGGGGCTGTTCACACAGACACATCTGGGCCAT 5021
 QY 481 CGGTGTGACAGCTTCTCATCTGTGAGCGCCGAGTACCCGGTGAGCAGCGAGGGTGACG 540
 DB 5022 CGGTGTGACAGCTTCTCATCTGTGAGCGCCGAGTACCCGGTGAGCAGCGAGGGTGACG 5081

QY 541 CGACCTGCAGAGAGACTGCGCAAGGGGCTGTGTCGGCTGAGCGGCTGCTACGGGGGCT 600
 DB 5082 CGACCTGCAGAGAGACTGCGCAAGGGGCTGTGTCGGCTGAGCGGCTGCTACGGGGGCT 5141
 QY 601 GTCGGGGGGCGCGCTTCCTTCGCGAGCACTGTGTCGCCCGCCAGCCCTGGGGCCGCC 660
 DB 5142 GTCGGGGGGCGCGCTTCCTTCGCGAGCACTGTGTCGCCCGCCAGCCCTGGGGCCGCC 5201
 QY 661 CTGCGCGTCCAGCTGCTGCACTGCGGCTGCGGCTAACAGAGAGGGCCCGGCTGTGTCGCG 720
 DB 5202 CTGCGCGTCCAGCTGCTGCACTGCGGCTAACAGAGAGGGCCCGGCTGTGTCGCG 5261
 QY 721 GCCCGCGCGCTGACCTTCCTACCGCGCGCTGCGGCTGCTGCGCAATCGGAGAA 780
 DB 5262 GCCCGCGCGCTGACCTTCCTACCGCGCGCTGCGGCTGCTGCGCAATCGGAGAA 5321
 QY 781 GAGAGCGCGCTGCTGTATACAGAGGACCAACTGCGCCAGGTGCGCTGAGGGTGGCCCG 840
 DB 5322 GAGAGCGCGCTGCTGTATACAGAGGACCAACTGCGCCAGGTGCGCTGAGGGTGGCCCG 5381
 QY 841 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGAGACGCGCAATAAAAAACGCCAGC 900
 DB 5382 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGAGACGCGCAATAAAAAACGCCAGC 5441
 QY 901 CGGGCGAGAAAAA 913
 DB 5442 CGGGCGAGAAAAA 5454

RESULT 8
 AAH46904
 ID AAH46904 standard; cDNA; 4936 BP.
 XX
 AC AAH46904;
 XX
 DT 25-SEP-2001 (first entry)
 XX
 DE cDNA encoding human protein kinase SGK145.
 XX
 KW Protein kinase; enzyme; cytostatic; nootropic; neuroprotective; human;
 KW antiparkinsonian; virucide; antibacterial; antifungal; antimigraine;
 KW analgesic; hypotensive; hypertensive; immunosuppressive; antiallergic;
 KW antipsoriatic; antirheumatic; antiarthritic; ophthalmological; anorectic;
 KW osteopathic; thrombolytic; antiarteriosclerotic; antiasthmatic;
 KW vasotropic; antidiabetic; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..4851
 FT /*tag= a
 XX
 EN WO200155356-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US002337.
 XX
 PR 25-JAN-2000; 2000US-0178078P.
 PR 31-JAN-2000; 2000US-0179364P.
 PR 17-FEB-2000; 2000US-0181173P.
 PR 17-MAR-2000; 2000US-0190162P.
 PR 29-MAR-2000; 2000US-0193404P.
 PR 13-NOV-2000; 2000US-0247013P.
 XX
 PA (SUGEN-) SUGEN INC.
 XX
 PI Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;
 XX WPI; 2001-476202/51.
 DR P-PSDB; AAB85504.
 XX
 PT Kinase polypeptides useful for treating cancers, Alzheimer's disease,

PT viral infections, diabetes, obesity, organ transplant rejection and
PT rheumatoid arthritis.
XX
PS Example 1; Page 204-205; 218pp; English.
XX
The invention provides human protein kinases and protein kinase-like
enzymes and polynucleotides encoding the polypeptides. The kinase
polypeptides and their modulators are useful for treating a disease or
disorder such as cancer, immune-related diseases, cardiovascular disease,
brain or neuronal-associated disease and metabolic disorders, including
cancers of tissues, cancers of hematopoietic origin, diseases of the
central nervous system, diseases of the peripheral nervous system,
Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic
lateral sclerosis, viral infections, infections caused by prions,
bacteria and fungi, ocular diseases, migraines, pain, sexual dysfunction,
mood disorders, attention disorders, cognition disorders, hypotension,
hyperextension, psychotic disorders, neurological disorders, dyskinesias,
metabolic disorders, and organ transplant rejection. They are also useful
for treating rhinitis, autoimmunity, atherosclerosis, psoriasis,
osteoarthritis, asthma, chronic inflammatory pelvic disease, chronic
inflammatory bowel disease, rheumatoid arthritis, metabolic disorders
such as diabetes, obesity, cardiovascular diseases such as reperfusion
injury, coronary thrombosis, clotting disorders and atherosclerosis,
ocular diseases such as glaucoma, retinopathy and macular degeneration,
psychiatric and neurological disorders such as anxiety, schizophrenia,
dementia, manic depression, etc. The polynucleotides are useful in gene
therapy techniques to treat the above mentioned disorders. Sequences
AAH46831-46922 represent human protein kinases encoding cDNA molecules
SQ Sequence 4936 BP; 923 A; 1695 C; 1524 G; 794 T; 0 U; 0 Other;

Query Match 93.3%; Score 906.4; DB 4; Length 4936;
Best Local Similarity 99.9%; Pred. NO. 1.2e-159;
Matches 907; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCGCTGGCGCCGACAGATCATCCCTACACCCCAAGGACAGACGAGT 60
DB 4029 CAGCGGGGGGGCGCTGGCGCCGACAGATCATCCCTACACCCCAAGGACAGACGAGT 4088
QY 61 GCTGCGGAATACGAGGCGCTCAAGGGCTGGCGCACCCGCACTGCGCCAGCTGCACG 120
DB 4089 GCTGCGGAATACGAGGCGCTCAAGGGCTGGCGCACCCGCACTGCGCCAGCTGCACG 4148
QY 121 AGCTACCTCAGCCCGCCGCGACCTGGTCTCATCTTGGAGCTGCTCTGGCGCCGAGCT 180
DB 4149 AGCTACCTCAGCCCGCCGCGACCTGGTCTCATCTTGGAGCTGCTCTGGCGCCGAGCT 4208
QY 181 GCTCCCTGCTGGCGGAGGGCTCTCTACTCAGAAATCCGAGTGAAGACTACCTGTG 240
DB 4209 GCTCCCTGCTGGCGGAGGGCTCTCTACTCAGAAATCCGAGTGAAGACTACCTGTG 4268
QY 241 GCAGATGTGAGTGCACCCAGTACTGCACACGACACATCTGACCTGGACCTGAG 300
DB 4269 GCAGATGTGAGTGCACCCAGTACTGCACACGACACATCTGACCTGGACCTGAG 4328
QY 301 GTCCGAGAACATGATCATCACCGAATACAACTGCTCAAGTCTGTGACCTGGGCAATGC 360
DB 4329 GTCCGAGAACATGATCATCACCGAATACAACTGCTCAAGTCTGTGACCTGGGCAATGC 4388
QY 361 ACAGAGCTTCAGCAGGAGAGGTGCTCCCTCAGCAAGTTCAGAGTACCTAGAGAC 420
DB 4389 ACAGAGCTTCAGCAGGAGAGGTGCTCCCTCAGCAAGTTCAGAGTACCTAGAGAC 4448
QY 421 CATGGCTCAGAGCTCTCGAGGGCCAGGGGCTCTCCACAGACAGACATCTGGGCCAT 480
DB 4449 CATGGCTCAGAGCTCTCGAGGGCCAGGGGCTCTCTCCACAGACAGACATCTGGGCCAT 4508
QY 481 CGGTGTGACAGCTTTCATCATCTGAGCGCCGAGTACCCGGTGAACAGCGAGGTGCACG 540
DB 4509 CGGTGTGACAGCTTTCATCATCTGAGCGCCGAGTACCCGGTGAACAGCGAGGTGCACG 4568
QY 541 CCACCTGCAGAGAGACTCCGCAAGGGCTGTCTCGGCTGAGCCCTGCTACGCGGGCT 600

DB 4569 CGACCTGCAGAGAGGACTGCGCAAGGGGTGTCTCCGGTGAAGCCCTGCTACGCGGGGT 4628
QY 601 GTCCGGGGGGCGCGCTGGCGCTTCTGCGCAGACACTGTGCGCCAGCCCTGGGGCGCGGC 660
DB 4629 GTCCGGGGGGCGCGCTGGCGCTTCTGCGCAGACACTGTGCGCCAGCCCTGGGGCGCGGC 4688
QY 661 CTGCGCGTCCAGTGCCTGCTGCACTGCGCGCTTAACAGAGAGGGCCCGGCTGTTCGGC 720
DB 4689 CTGCGCGTCCAGTGCCTGCTGCACTGCGCGCTTAACAGAGAGGGCCCGGCTGTTCGGC 4748
QY 721 GCCCGCGCGCTGACCTTCCCTACCGCGGCTGCGCGCTTCTGCGCAATTCGGGAGAA 780
DB 4749 GCCCGCGCGCTGACCTTCCCTACCGCGGCTGCGCGCTTCTGCGCAATTCGGGAGAA 4808
QY 781 GAGACGCGCGCTGTGTACAGAGGACAACTGCGCCAGGTGCGCTGAGGGTTCGCCCCG 840
DB 4809 GAGACGCGCGCTGTGTACAGAGGACAACTGCGCCAGGTGCGCTGAGGGTTCGCCCCG 4868
QY 841 GCCACACCTTGTCTCCCGCTGGGGGTGCTGCGAGCGCCCAATAAAGCGCCAGC 900
DB 4869 GCCACACCTTGTCTCCCGCTGGGGGTGCTGCGAGCGCCCAATAAAGCGCCAGC 4928
QY 901 CGGGCGAG 908
DB 4929 CGGGCGAG 4936
RESULT 9
AAC62286
ID AAC62286 standard; cDNA; 5007 BP.
XX
AC AAC62286;
XX
DT 19-MAR-2001 (first entry)
XX
DE cDNA encoding a full length human signal transduction polypeptide.
XX
KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
KW congestive heart failure; dilated congestive cardiomyopathy;
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
KW mitral valve disease; aortic valve disease; tricuspid valve disease;
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;
KW atherosclerosis; cardiac tumour; microbial infection; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 97..4929
FT /tag= a
FT /product= "signal transduction polypeptide H19G5"
XX
PN WO200063381-A1.
XX
PD 26-OCT-2000.
XX
PF 11-APR-2000; 2000WO-US009488.
XX
PR 16-APR-1999; 99US-0129553P.
XX
PA (SCIO-) SCIOS INC.
XX
PI Zeng W, Stanton L, Kong H;
XX
PI WPI; 2001-007013/01.
XX
DR P-PSDB; AAB30568.
XX
PT Novel h19G5 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
PT cardiac diseases, and additional mediators of signal transduction.
XX
PS Claim 4; Page 59-61; 81pp; English.
XX
CC The present sequence encodes a human protein with putative function in

transduction or the expression of a protein capable of acting as a donor or acceptor molecule of a phosphate group. The monoclonal antibodies can be used as probes for detecting discrete antigens expressed by tissue or cell samples, and therefore used in humans for localization and monitoring of microbial infection

XX SQ Sequence 7928 BP; 1535 A; 2644 C; 2488 G; 1261 T; 0 U; 0 Other;

Query Match 93.2%; Score 905.4; DB 5; Length 7928;
Best Local Similarity 99.9%; Pred. No. 2e-159;
Matches 906; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGGCGAGATCATCCCTACCCAGGACAGACAGCAGT 60
DB 7022 CAGCGGGGGGGCTGGCGGCGAGATCATCCCTACCCAGGACAGACAGCAGT 7081

QY 61 GCTGGCGCAATACAGAGGCGCTCAAGGGCTGGCGGACCCGACCTGGCCAGTGCACGC 120
DB 7082 GCTGGCGCAATACAGAGGCGCTCAAGGGCTGGCGGACCCGACCTGGCCAGTGCACGC 7141

QY 121 AGCTTACCTCAGCCCGCGGACCTGGTGTCTATCTTGGAGTGTGTCTGGCGCCGAGCT 180
DB 7142 AGCTTACCTCAGCCCGCGGACCTGGTGTCTATCTTGGAGTGTGTCTGGCGCCGAGCT 7201

QY 181 GCTCCCTCTGCTGGCGGAGAGGCTCTCTACTCAGATCCAGGTGAAGACTACCTGTG 240
DB 7202 GCTCCCTCTGCTGGCGGAGAGGCTCTCTACTCAGATCCAGGTGAAGACTACCTGTG 7261

QY 241 GCAGATGTGAGTGCACCCAGTACCTGCACAAACAGACATCTCTGCACCTGGACCTGAG 300
DB 7262 GCAGATGTGAGTGCACCCAGTACCTGCACAAACAGACATCTCTGCACCTGGACCTGAG 7321

QY 301 GTCGAGACATGATCATCCGATACAACTGCTCAAGTCTGGACCTGGCCAAATGC 360
DB 7322 GTCGAGACATGATCATCCGATACAACTGCTCAAGTCTGGACCTGGCCAAATGC 7381

QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGCTCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420
DB 7382 ACAGAGCTCAGCCAGGAGAGGTGCTGCTCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 7441

QY 421 CATGCTCAGAGCTCTTGAGGGCCAGGGGCTGTTCCAGACAGACATCTGGGCCAT 480
DB 7442 CATGCTCAGAGCTCTTGAGGGCCAGGGGCTGTTCCAGACAGACATCTGGGCCAT 7501

QY 481 CGGTGTGACAGCTTTCATCATCTGAGCGCCAGTACCCGGTGAAGCGAGGGTGCACG 540
DB 7502 CGGTGTGACAGCTTTCATCATCTGAGCGCCAGTACCCGGTGAAGCGAGGGTGCACG 7561

QY 541 CGACCTGACAGAGACTGCGCAAGGGCTGCTCGGCTGAGCCGCTCTACCGGGGCT 600
DB 7562 CGACCTGACAGAGACTGCGCAAGGGCTGCTCGGCTGAGCCGCTCTACCGGGGCT 7621

QY 601 GTCGGGGGGCGGCTGCTTCTCGGACGACTCTGCGCCAGCCCTGGGGCGGCGC 660
DB 7622 GTCGGGGGGCGGCTGCTTCTCGGACGACTCTGCGCCAGCCCTGGGGCGGCGC 7681

QY 661 CTGCGCGTCCAGTGCCTGAGTGCCTGCTGCTAAACAGAGAGGGCCGCGCTGTTCGCG 720
DB 7682 CTGCGCGTCCAGTGCCTGAGTGCCTGCTGCTAAACAGAGAGGGCCGCGCTGTTCGCG 7741

QY 721 GCGCGGCGGCTGACTTCTTCTACCGCGGCTGCGCTCTCTGCTGCGCAATCGCGAA 780
DB 7742 GCGCGGCGGCTGACTTCTTCTACCGCGGCTGCGCTCTCTGCTGCGCAATCGCGAA 7801

QY 781 GAGAGCGGCTGCTGTACAGAGGCACAACCTGGCCCGAGGTCGCTGAGGGTGCCTCCCG 840
DB 7802 GAGAGCGGCTGCTGTACAGAGGCACAACCTGGCCCGAGGTCGCTGAGGGTGCCTCCCG 7861

QY 841 GGCACACCTTGGTCTTCCCGTGGGGTGTGCTGACAGCGCGCCCAATAAAGCCCGCAG 900
DB 7862 GGCACACCTTGGTCTTCCCGTGGGGTGTGCTGACAGCGCGCCCAATAAAGCCCGCAGC 9621

QY 901 CGGGCGA 907

Db 7922 CGGGCGA 7928

RESULT 11

ABN76255

ID AEN76255 standard; cDNA; 225 BP.

XX AC AEN76255;

XX DT 08-JUL-2002 (first entry)

XX DE Human kinase-like ORF1202 cDNA, SEQ ID NO:2403.

XX KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;

XX KW disease monitoring; cytokine; cell proliferation; cell differentiation;

XX KW immune modulation; haematopoiesis regulation; tissue growth;

XX KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;

XX KW thrombolytic; tumour inhibition; bodily characteristic; fertility;

XX KW behaviour; cancer; proliferative disorder; neurological disorder;

XX KW cardiovascular disease; immune system disorder; organ transplantation;

XX KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;

XX KW hypothyroidism; cholesterol ester storage disease; infection; vulnary;

XX KW vasotrophic; antiparietic; antidiabetic; cytosolic; neurotropic;

XX KW neuroprotective; antithrombotic; anticoagulant; thrombolytic;

XX KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;

XX KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.

XX OS Homo sapiens.

XX XX WO200190366-A2.

XX XX 29-NOV-2001.

XX XX 24-MAY-2001; 2001WO-US017076.

XX XX 24-MAY-2000; 2000US-0206690P.

XX XX (CURA-) CURAGEN CORP.

XX XX Leach MD, Shinkets RA;

XX XX WPI; 2002-106200/14.

XX XX P-PSDB; ABP32229.

XX XX Novel human polypeptides and polynucleotides useful for diagnosing,

XX XX preventing and treating cardiovascular disease, neurodegenerative,

XX XX hyperproliferative disorders and disorders related to organ

XX XX transplantation.

XX XX Claim 1; Page 863; 2508pp; English.

XX XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins

XX XX designated ORF (open reading frame) 1-4534, and sequences ABN75054-

XX XX ABN79587 represent cDNAs encoding them. The invention also encompasses

XX XX polypeptides at least 80% identical to the ORF1-ORF4534 (collectively

XX XX referred to as ORFX) proteins, polynucleotides at least 85% identical to

XX XX the ORFX nucleic acid sequences, vectors and host cells comprising ORFX

XX XX polynucleotides, the recombinant production of ORFX proteins, antibodies

XX XX specific for ORFX proteins, methods of detecting ORFX polynucleotides and

XX XX polypeptides, methods of screening for modulators of ORFX expression or

XX XX activity, and methods of screening individuals for a predisposition to an

XX XX ORFX-associated disorder. The ORFX proteins of the invention have a wide

XX XX range of biological activities, such as cytokine, cell proliferation,

XX XX cell differentiation, immune modulation, haematopoiesis regulation,

XX XX tissue growth, angiogenesis, activin or inhibin activity, chemotactic/

XX XX chemokinetic activity, haemostatic activity, thrombolytic activity,

XX XX receptor/ligand, antiinflammatory activity, tumour inhibition activity,

XX XX and antiinfective activity, and may also be involved in the determination

XX XX of bodily characteristics, fertility and behaviour. ORFX proteins, c

XX XX nucleic acids and antibodies may be used in the treatment of cancers,

XX XX other proliferative disorders such as psoriasis and benign tumours,

XX XX neurological disorders such as epilepsy and Alzheimer's disease,

CC cardiovascular diseases, immune system disorders, disorders related to
 CC organ transplantation, disorders of tissue growth and regeneration,
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
 CC storage disease, and infectious diseases caused by viral, bacterial,
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a
 CC source of primers and probes, in the detection of ORFX genomic sequences
 CC or transcripts, in the identification and cloning of homologous
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases
 XX
 SQ Sequence 225 BP; 54 A; 65 C; 69 G; 37 T; 0 U; 0 Other;

Query Match 23.0%; Score 223.4; DB 6; Length 225;

Best Local Similarity 99.6%; Pred. No. 1.4e-32;
 Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 314 ATCATCACCGAATACAACTGCTCAAGTCTGACCTGGCAATGCACAGAGCTCAGC 373
 DB |||||
 QY 374 CAGGAGAGGTGCTGCTCTCAGACAAAGTTCAAGACTACCTAGACACATGGTCTCCAGAG 433
 DB |||||
 QY 61 CAGGAGAGGTGCTGCTCTCAGACAAAGTTCAAGACTACCTAGACACATGGTCTCCAGAG 120
 DB |||||
 QY 434 CTCCTGGAGGGCCAGGGGGCTGTTCCACAGACACATCTGGGCTCATGGTGTGACAGCC 493
 DB |||||
 QY 121 CTCCTGGAGGGCCAGGGGGCTGTTCCACAGACACATCTGGGCTCATGGTGTGACAGCC 180
 DB |||||
 QY 494 TTCATCATGCTGAGCGCCAGTACCCGGTGTGACAGCGAGGTGCA 538
 DB |||||
 QY 181 TTCATCATGCTGAGCGCCAGTACCCGGTGTGACAGCGAGGTGCA 225
 DB |||||

RESULT 12

AAAF4662

ID AAF44662 standard; cDNA; 7710 BP.

AC AAF44662;

DT 27-MAR-2001 (first entry)

DE Novel protein kinase cDNA, SEQ ID NO: 42.

XX Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
 KW immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
 KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;
 KW immune disorder; cardiovascular disease; neurodegenerative disease;
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.

OS Homo sapiens.

PN WO2000073469-A2.

PP 07-DEC-2000.

XX 26-MAY-2000; 2000WO-US014842.

XX 28-MAY-1999; 99US-0136503P.

XX (SUGEN-) SUGEN INC.

XX Plowman GD, Martinez R, Whyte D, Sudersanam S;

XX WPI; 2001-032161/04.

DR P-PSDB; AAB65635.

XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and
 PT treating immune-related diseases and disorders, cardiovascular disease,

PT neurodegenerative diseases and/or cancers.

XX Example 1; Fig 2; 310pp; English.

XX The present sequence encodes a novel protein kinase. The nucleic acids
 CC and the protein kinases they encode may be used in the treatment and
 CC diagnosis of diseases associated with inappropriate kinase expression
 CC such as immune-related diseases and disorders, cardiovascular disease,
 CC neurodegenerative diseases and/or cancers. The nucleic acids and
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays. The kinase polypeptides may be used as antigens in the production
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies
 CC and kinase antagonists may also be used to down regulate kinase
 CC expression and activity. Diseases related to kinase expression and
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
 CC disorders, complications of organ transplantation, myocardial infarction,
 CC immune disorders, cardiomyopathies, strokes, renal failure, oxidative-
 CC stress related disorders, chronic inflammatory bowel disease, chronic
 CC inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis,
 CC psoriasis, rhinitis, autoimmunity, diabetes, cancers and reproductive
 CC disorders

SQ Sequence 7710 BP; 1446 A; 2534 C; 2447 G; 1283 T; 0 U; 0 Other;

Query Match 21.2%; Score 206; DB 4; Length 7710;

Best Local Similarity 53.3%; Pred. No. 4e-29;

Matches 434; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

QY 1 CAGCGGGGGCGGTGGCGCCCAAGATCATCCCTTACCACCCCAAGGACAGAGCAGT 60
 DB |||||
 QY 6018 CAGCGGGCGAAGCTTCGTGGGCCAAGATCGTCCCTATCTGCCAGGCAAGCCCGGGT 6077
 DB |||||
 QY 61 GTTCGCGGAATACGAGGCGCTCAAGGGCTGGCGCACCCGACCTGGCCCACTGGCAGC 120
 DB |||||
 QY 6078 CCGTCAGGAGTACGAGGTGCTGGGAGCCCTGCACCACGAGCGGATCATGTCCTGCACGA 6137
 DB |||||
 QY 121 AGCTTACTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTGCTCTGGGCCGAGCT 180
 DB |||||
 QY 6138 GGCCTACATACCCCTCGGTACCTCGTGTCTCATCTGTGAGAGCTGTGGCAACCGGGAAT 6197
 DB |||||
 QY 181 GCTCCCTCGCTGGCGGAGAGGGCTCCTTACTCAGATCCAGAGTGAAGGACTACCTGTG 240
 DB |||||
 QY 6198 CCTCTGGGCTCAGTGACAGGTTCGGGTATCTTGAGGATGACGTGGCCACTTACATGT 6257
 DB |||||
 QY 241 GCAGATGTTGAGTGCCACCCAGTACCTGCACACAGGACATCTCTGCACCTGGACCTGAG 300
 DB |||||
 QY 6258 GCAGTGTCTAAGGCTGGACTACCTCCAGGGCCACCCAGCTGTCCACCTAGACATCAA 6317
 DB |||||
 QY 301 GTCGAGAACATGATCATCACCAGATACAACTGCTCAAGTCTGTGACCTGGGCAATGC 360
 DB |||||
 QY 6318 GCCAGAACCTGTGCTGGCCCTGACATGCCCTCAAGATTGTGGACTTTGGCAGTGC 6377
 DB |||||
 QY 361 ACAGAGCTCTAGCCAGGAGAAGGTGCTCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420
 DB |||||
 QY 6378 CCAGCCCTACAAACCCCGAGGCGCTTAGGCGCCCTTGGCCACCGCACGCTGGAGTT 6437
 DB |||||
 QY 421 CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCAAT 480
 DB |||||
 QY 6438 CATGGCTCCGAGATGTTGAAGGAGAAACCCATCGGCTCTGCCACGACATCTGGGAGC 6497
 DB |||||
 QY 481 CGGTGTGACAGCTTTCATCATGCTGAGCGCCGAGTACCCGGTGTGAGCAGCGAGGTGCAQC 540
 DB |||||
 QY 6498 GGGTGTGCTCACTTACATTAATGCTCAGTGAGCGCTCCCGGTTCTATGAGCCAGACCCCA 6557
 DB |||||
 QY 541 CGACCTGCAGAGAGACTGCCAGGGGCTGGTCCGGCTGAGCGCTGCTACCGGGGCT 600
 DB |||||
 QY 6558 GGAACGAGAGGCTCGAATGTGGGGGCGCGCTTTGATGCCTTCAGGTGTACCCCAATAC 6617
 DB |||||
 QY 601 GTCCGGGGCGCGGTGGCTTCTCTGCGGAGCACTCTGTGCGCCAGCCCTGGGGCCGCGCC 660
 DB |||||
 QY 6618 ATCCAGAGCGCCACCTCTTCTTGGGAAAGGTTCTCTCTGTATCATCTCTGGAGCGCGCC 6677
 DB |||||
 QY 661 CTGCGCTGCCAGTGCCTGTGAGTCCCGTGGCTAACAGAGAGGCGCGCGCTTTCGCG 720
 DB |||||

Db 6678 CTCCTGTGAGGAGCTGCTGGCCCAACCATGGTTGACGAGCCCTACTGTATGAAGTGGC 6737
Qy 721 GCCCGCGCGCTGACCTTCCCTACCGCGCGGCTGGGCTCTTCGTGGCAATCGCGAGAA 780
Db 6738 CGGCAGACGCTCACCITTCACCAACCGGCTCAAGAGTTCCTGGCGAGCAGCGGCG 6797
Qy 781 GAGAGCGCGCTGCTGTACAGAGGCGCAACCTG 814
Db 6798 GCGCGGCGCTGAGGCTGCCACCGCCCAAGGTG 6831

RESULT 13

AAD30565 standard; cDNA; 7789 BP.

XX AAD30565;

XX 21-MAY-2002 (first entry)

XX Human kinase polypeptide (PKIN-18) cDNA.

Kw Human; kinase polypeptide; PKIN-18; gene therapy; Addison's disease;
Kw leukaemia; immune disorder; lymphoma; melanoma; developmental disorder;
Kw acquired immunodeficiency syndrome; AIDS; allergy; anaemia; hypertension;
Kw asthma; Crohn's disease; rheumatoid arthritis; bursitis; atherosclerosis;
Kw cirrhosis; hepatitis; psoriasis; Cushing's syndrome; cytostatic; cancer;
Kw cholestasis; cardiac; cardiovascular disorder; Niemann-Pick's disease;
Kw lipid disorder; fatty liver; Gaucher's disease; myocardial infarction;
Kw drug screening; transgenic animal; antiinflammatory; hepatotropic;
Kw hypotensive; anti-HIV; enzyme; ss.

XX Homo sapiens.

XX WO200208399-A2.

XX 31-JAN-2002.

XX 20-JUL-2001; 2001WO-US023092.

XX 21-JUL-2000; 2000US-0220038P.

XX 28-JUL-2000; 2000US-0222112P.

XX 04-AUG-2000; 2000US-0222831P.

XX 11-AUG-2000; 2000US-0224729P.

XX (INCY-) INCYTE GENOMICS INC.

XX (THOR/) THORNTON M.

XX Thornton M, Yue H, Khan FA, Gururajan R, Hafalia AJA, Wallia NK;

XX Patterson C, Rankumar J, Gandhi AR, Policky JL, Baughn MR;

XX Tribouley CM, Bandman O, Nguyen DB, Lu Y, Burford N, Lal P, Ding L;

XX Yao MG, Elliott VS, Recipon SA, Kearney L, Lu DAM, Greenwald SR;

XX Tang YT, Xu Y, Walsh RT, Gietzen KJ, Yang J, Hillman JL;

XX WPI; 2002-206083/26.

XX New human kinase polypeptide, useful in diagnosis, prevention and

XX treatment of cancer, immune disorder, growth and developmental disorder,

XX cardiovascular disorder and lipid disorder.

XX Claim 5; Page 191-193; 196pp; English.

XX The present invention relates to an isolated human kinase polypeptide
XX (PKIN) or a biologically active/immunogenic fragment of PKIN. PKIN is
XX useful for diagnosing, treating and preventing cancer (e.g., leukaemia,
XX lymphoma, melanoma), an immune disorder (e.g., acquired immunodeficiency
XX syndrome (AIDS), Addison's disease, allergy, anaemia, asthma, Crohn's
XX disease, rheumatoid arthritis), a growth and developmental disorder (e.g.
XX bursitis, cirrhosis, hepatitis, psoriasis, Cushing's syndrome), a
XX cardiovascular disorder (e.g., atherosclerosis, hypertension, myocardial
XX infarction), and a lipid disorder (e.g., fatty liver, cholestasis,
XX Gaucher's disease, Niemann-Pick's disease). PKIN is useful in a number of
XX drug screening techniques and to analyse the proteome of a tissue or cell

CC type. PKIN is useful for creating knockin humanised animals or transgenic
CC animals to model human diseases, in somatic or germline gene therapy, to
CC generate a transcript image of a tissue or cell type, for detecting
CC differences in the chromosomal location due to translocation, inversion,
CC etc., among normal, carrier or affected individuals, and as hybridisation
CC probes for mapping naturally occurring genomic sequences. PKIN is useful
CC in southern or northern analysis, dot blot or other membrane-based
CC technologies, in PCR technologies, in dipstick, pin, multifomat enzyme
CC linked immunosorbent (ELISA)-like assays and in microarrays utilising
CC fluids or tissues from patients to detect altered PKIN expression. The
CC present sequence is human PKIN-18 cDNA. Note: This sequence is said to
CC encode PKIN-18 referred as SEQ ID NO:18 (AAE19160). However this does not
CC appear to be the case

XX Sequence 7789 BP; 1469 A; 2558 C; 2470 G; 1292 T; 0 U; 0 Other;

SQ Query Match 21.1%; Score 204.4; DB 6; Length 7789;

Best Local Similarity 53.2%; Pred. No. 8e-29;

Matches 433; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

Qy 1 CAGCGGGCGGCGCTGGCGCAAGATCATCCCTACCACCCAGGACAAAGACAGCACT 60
Db 6080 CACGGGGCAAGATTCTGTCGCCAAGATCTGTCCTATGTGCGAGGGCAAGCGGGGT 6139
Qy 61 GTTCGCGCAATACGAGGCGCTCAAGGGCTGCGCCACCGACCTGGCCAGCTGCACGC 120
Db 6140 CTTGCAGGAGTACGAGTGTCTGGGACCTGCACACGAGCGATCATGTCCCTGCACGA 6199
Qy 121 AGCTACCTCAGCCCGGCGACCTGTGCTCATCTTGGAGCTGTCTGGGCCCGAGCT 180
Db 6200 GGCTCATACACCCCTCGGTACCTCTGCTCATCTGAGAGCTGTGGCAACCGGAAC 6259
Qy 181 GCTCCCTCTGCTGGCGGAGAGGCGCTCTTACTCAGAAATCCGAGGTGAAGACTTACCTGTG 240
Db 6260 CCTCTGCGGCTCAGTACAGGTTCCGGTATCTGAGGATGACGTGCCACATACATGGT 6319
Qy 241 GCAGATGTTAGTGCACCCAGTACTGTGCACAAACGAGACATCTGACCTGAGCACTGAG 300
Db 6320 GCAGCTGCTACAGGCGCTGAGTACTTCCACGCGCCACCGTCTCCACTGACATCAA 6379
Qy 301 GTCGAGACATGATCATCACGAAATACAACTGCTCAAGGTCTGTCGAGCTGGGCAATGC 360
Db 6380 GCAGACACCTGCTGCTGGCCCTGACAAATGCCCTCAGATGTGAGCTTTGGCAGTGC 6439
Qy 361 ACAGAGCCTCAGCCAGGAGAGGTGTGCTCCTCAGACAAAGTTCAAGGATCTCTAGAGAC 420
Db 6440 CCAGCGCTACACCCCGCGGCGCTTAGGGCCCTTGGCCACCGCGGACGCTGGAGTT 6499
Qy 421 CATGGCTCCAGAGCTCTCGAGGCGCCAGGGGGTGTTCACAGACAGACATCTGGGCCAT 480
Db 6500 CATGGCTCCGAGATGTTGAGGGAGAACCCATCGGCTGTGCCACGAGACATCTGGGGAGC 5559
Qy 481 CGGTGTACAGCCTTTCATCATGTGAGCGCGGAGTACCCGGTGAAGCAGCGAGGTGCACG 540
Db 6560 GGGTGTGCTCACCTTACATATGCTCAGTGAAGCTCCCGTTCTATGAGCCAGACCCCA 6619
Qy 541 GCACCTGCAGAGGAGTGGCGAAGGGGTGCTCGGCTGAGCGGCTGTACCGGGGCT 600
Db 6620 GGAACCGAGGCTCGGATTTGTGGGGGGCGGCTTGTATGCTTCCAGCTGTACCCCAATAC 6679
Qy 601 GTCGCGGGGCGCGTGGCGCTTCTCGCGCAGACATCTGTGTGCGCCAGCCTCTGGGCGCGGCC 660
Db 6680 ATCCAGAGCGCCACCTCTTCTTGGAAAGGTTCTCTGTGACATCTCTGGAGCGGCC 6739
Qy 661 CTCGGCGTCAAGTGTGCTGAGTCCCGTGGCTGTAAAGAGGAGGGCGCGCTGTTCGCG 720
Db 6740 CTCCTGCAGGACTGCTGCGCCACCCATGTTGACAGGACGCGCTACTCTATGAAGTGGC 6799
Qy 721 GCCCGCGCGCTGACCTTCCCTACCGCGGCTGCGCTCTTCGTGGCAATCGCGAGAA 780
Db 6800 CCGCGAGACCTCACCTTACACCAACCGGCTCAGGAGTTCTTGGCGGAGCAGCGCGG 6859
Qy 781 GAGACGCGCGCTGCTGTACAAAGAGGCACAACTG 814

| | | | | | | | |
|-----------|---|---|---------|-----------|---|--|------|
| DB | 6860 | GGCGGGCTGAGGTGCGCCCGCCACAAGGTG | 6893 | QY | 61 | GCTGCGGAATACGAGGCGCTCAAGGGCTGGCGCCACCGGACCTTGGCCCGCTGCACGC | 120 |
| DB | | | | DB | 9024 | CCTGACAGGAGTACGAGGTGCTGCGGACCTGTCACACGAGCGGATCATGTCTCTGCACGA | 9083 |
| RESULT 14 | | | | QY | 121 | AGCTACCTCAGCCCCCGGACCTGCTCATCTTGGAGCTGTGCTCTTGGGCCCGAGCT | 180 |
| ID | ADB79958 | | | DB | 9084 | GGCTACATCACCCCTCGGTACTCTGCTCATCTGAGAGCTGTGGCAACCGGAACT | 9143 |
| XX | ADB79958 standard, cDNA; 9807 BP. | | | QY | 181 | GCTCCCTGCTGGCGGAGAGGCTCTCTACAGAAATCCGAGGTGAAGGACTACCTGTG | 240 |
| XX | AC | | | DB | 9144 | CCTCTGTGGGCTCAGTGACAGGTTCGGGTATTCTGAGGATGACGTGGCCACTTACATGTT | 9203 |
| XX | 04-DEC-2003 (first entry) | | | QY | 241 | GCAGATGTTGAGTGCCACCCAGTACCTGCACAAACAGCACATCTCTGACCTGGAGCCTGAG | 300 |
| DE | Human kinase protein encoding cDNA SEQ ID NO:1. | | | DB | 9204 | GCAGCTGTCTACAGGCTGGGACTACTCCAGCGGCACACAGTCTCCACCTAGACATCAA | 9263 |
| XX | human; kinase; enzyme; chromosome 2; cytostatic; gene therapy; | | | QY | 301 | GTCCGAGAACATGATCATCACGGAATACAACCTGCTCAAGGTCTGAGGCTGGGCAATGC | 360 |
| KW | brain anaplastic oligodendroglioma; lung carcinoma; | | | DB | 9264 | GCAGACAACTGCTGCTGGCCCTGACAAATGCCCTCAAGATTGTGGAATTTGGCAGTGC | 9323 |
| KW | soft tissue leiomyosarcoma; ovary tumour; germ cell tumour; gene; ss. | | | QY | 361 | ACAGAGCCTCAGCCAGGAGAGGTGCTGCTCCCTCAGACAAAGTTCAAGGACTACCTAGACAC | 420 |
| OS | Homo sapiens. | | | DB | 9324 | CCAGCCTTACAAACCCAGGCGCTTAGGCCCTTGGCCACCGCACGGGACGCTGGAGTT | 9383 |
| XX | | | | QY | 421 | CATGCTCCAGAGTCTCTGGAGGCGCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT | 480 |
| PH | Key | Location/Qualifiers | | DB | 9384 | CATGCTCCGAGATGCTGAAGGAGAACCCATCGCTCTGCCACGGACATCTGGGGAGC | 9443 |
| FT | 1..9807 | /*tag= a | | QY | 481 | CGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCCGCTGAGCAGGAGGTGACG | 540 |
| FT | CDS | /product= "kinase protein" | | DB | 9444 | GGGTGTGCTCACTTACATTAATGCTCAGTGGAGCGTCCCGCTTCTATGAGCCAGACCCGCCA | 9503 |
| FT | | | | QY | 541 | CGACCTGCAGAGAGGACTGCGCAAGGGCTGGTCCGGCTGAGCGGCTGCTACGCGGGGCT | 600 |
| XX | WO2003076577-A2. | | | DB | 9504 | GGAAACGGAGGCTGGATTGTGGGGCGCGTTTGATGCTTCCAGCTGTACCCCAATAC | 9563 |
| PD | 18-SEP-2003. | | | QY | 601 | GTCCGGGGCGCGTGGGCTTCTGCGCAGCACTCTGTGCGCCGACGCTTGGGGCGCGCC | 660 |
| XX | 05-MAR-2003; 2003WO-US006666. | | | DB | 9564 | ATCCGAGAGCGCCACCCCTCTTCTTGGGAAAGGTTCTCTCTATACATCCCTGAGCGCGCC | 9623 |
| XX | 05-MAR-2002; 2002US-0361339P. | | | QY | 661 | CTGCGCGTCCAGCTGCTGCTGAGTGGCTGCTTAACAGAGGAGGCGCGCGCTGTTCGCG | 720 |
| PA | (APPL-) APPLERA CORP. | | | DB | 9624 | CTCCCTGCAGGACTGCTTGGCCCGCCCATGTTGCAGGACGCTTACCTGATGAGAGTGG | 9683 |
| PI | Yan C, Gan W; | | | QY | 721 | GCCCGCGCGGTGACCTTCCCTACGCGCGGCTGGGCTCTTCTGTCGGCAATCGCGGAA | 780 |
| DR | WPI; 2003-722329/68. | | | DB | 9684 | CGCCGAGACGCTACCTTACACCAACCGGCTCAAGGAGTTCTTGGCGGAGCAGCGCGG | 9743 |
| DR | P-FSDB; ADB79959. | | | QY | 781 | GAGAGCGCGCTGCTGTGTACAAGAGGACAAACCTG | 814 |
| XX | New peptides related to kinase protein subfamily useful for treating | | | DB | 9744 | GGCGCGGCTGAGGCTGCCACCCGCCACAGGTG | 9777 |
| PT | disorders associated with abnormal expression of kinase protein in | | | RESULT 15 | | | |
| PT | testis, nervous tissue, fetal, lung, ovary tumor tissue. | | | AD99127 | | | |
| XX | Claim 4; Fig 1A-C; 86pp; English. | | | ID | ADC99127 standard; cDNA; 10662 BP. | | |
| CC | The present sequence encodes a human kinase protein (I), which is located | | | XX | ADC99127; | | |
| CC | to chromosome 2. The present invention also describes an allelic variant | | | AC | ADC99127; | | |
| CC | or orthologue of (I). (I) has cytostatic activity, and can be used in | | | XX | 01-JAN-2004 (first entry) | | |
| CC | gene therapy. (I) can be used in substantial and specific assays related | | | DT | Human KPP cDNA - SEQ ID 80. | | |
| CC | to functional information of the protein sequence, to raise antibodies or | | | XX | anti-HIV; anti-allergic; anti-inflammatory; antianemic; antiparkinsonian; | | |
| CC | to elicit immune responses, as reagents in assays to determine the levels | | | XX | neurotropic; anticonvulsant; antiarteriosclerotic; antiasthmatic; | | |
| CC | of protein in biological fluids. (I) can be used in drug screening assays | | | XX | immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological; | | |
| CC | for identifying agents that are useful in treating disorders associated | | | XX | antidiabetic; nephrotropic; antitumor; thyromimetic; neuroprotective; | | |
| CC | with the absence of, inappropriate, or unwanted expression of kinase | | | XX | osteopathic; antiarthritic; antiparasitic; antihelminthic; antiprosperic; | | |
| CC | protein in testis, nervous tissue, foetal, lung, brain anaplastic | | | XX | uroepathic; opthoalmologic; antirheumatic; haemostatic; antibacterial; | | |
| CC | oligodendroglioma, lung carcinoma tissue, soft tissue leiomyosarcoma, | | | XX | virucide; protozoacide; fungicide; kinase; phosphatase; KPP; | | |
| CC | ovary tumour tissue, or germ cell tumour tissue. The protein and nucleic | | | XX | cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; | | |
| CC | acid sequences of (I) are useful as models for the development of human | | | XX | cancer; developmental; mental retardation; neurologic; Alzheimer's | | |
| CC | therapeutic targets, in the identification of therapeutic proteins and as | | | XX | disease; Parkinson's; autoimmune; inflammatory; Crohn's; | | |
| CC | targets for the development of human therapeutic agents that modulate | | | | | | |
| CC | protease activity in cells and tissues that express the kinase peptide. | | | | | | |
| XX | | | | | | | |
| XX | Sequence 9807 BP; 1721 A; 3370 C; 3208 G; 1508 T; 0 U; 0 Other; | | | | | | |
| XX | Query Match | 21.18; Score 204.4; DB 9; Length 9807; | | | | | |
| XX | Best Local Similarity | 53.2%; Pred. No. 8.3e-29; | | | | | |
| XX | Matches 433; Conservative | 0; Mismatches 381; Indels | 0; Gaps | 0; | | | |
| QY | 1 | CAGCGGGCGGGCTGGCGCCGACGATCATCCCTACACCCCAAGGACAAAGACAGCT | 60 | | | | |
| DB | 8964 | CACGGGCGGAACGTTCTGTCGCAAGATCGTGCCCTATGCTGCGAGGCGGAGCGGGGT | 9023 | | | | |

KW diabetes mellitus; viral; bacterial; fungal; parasitic; protozoan;
KW helminthic infection; transgenic; gene therapy; human; ss; gene.
XX Homo sapiens.
XX WO2003033680-A2.
XX 24-APR-2003.
XX 17-OCT-2002; 2002WO-US033723.
XX 19-OCT-2001; 2001US-0345474P.
XX 02-NOV-2001; 2001US-0343910P.
XX 13-NOV-2001; 2001US-0333098P.
XX 16-NOV-2001; 2001US-0332424P.
XX 30-NOV-2001; 2001US-0334288P.
XX (INCY-) INCYTE GENOMICS INC.
XX Bandman O, Baughn MR, Becha SD, Borowsky ML, Duggan BM;
XX Emerling BM, Forsythe IJ, Gandhi AR, Gorvad AE, Griffin JA;
XX Gururajan R, Hafalia AJA, Khan FA, Lal PG, Lee EA, Lee SY;
XX Lindquist EA, Lu DAM, Lu Y, Marquis JP, Nguyen DB, Arvizu CS;
XX Ramkumar J, Recipon SA, Richardson TW, Swarnakar A, Tang YT;
XX Thornton MB, Tran UK, Chawla NK, Warren BA, Yang J, Yao MG, Yue H;
XX Zebarjadian Y;
XX WPI; 2003-403214/38.
XX P-PSDB; ADC99075.
XX New human kinases and phosphatases and polynucleotides, useful for
XX diagnosing, treating or preventing autoimmune or inflammatory disorders
XX (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,
XX cancer or hepatitis.
XX Claim 5; SEQ ID NO 80; 424pp; English.
XX The invention relates to a novel isolated polypeptide which is a human
XX kinase and phosphatase (KPP). The KPP polypeptides, polynucleotides,
XX agonists and antagonists are useful for diagnosing, treating or
XX preventing cell proliferative disorders such as atherosclerosis,
XX cirrhosis, hepatitis and cancer, developmental disorders e.g. mental
XX retardation, neurological disorders including Alzheimer's disease and
XX Parkinson's disease, autoimmune and inflammatory disorders such as
XX Crohn's disease and diabetes mellitus and finally, viral, bacterial, the
XX fungal, parasitic, protozoan or helminthic infections. Furthermore, the
XX polynucleotides encoding KPP may be useful for creating transgenic
XX animals to model human disease, as well as during gene therapy
XX procedures. The current sequence is that of the human KPP cDNA of the
XX invention.

SQ Sequence 10662 BP; 1933 A; 3587 C; 3486 G; 1656 T; 0 U; 0 Other;
Query Match 21.1%; Score 204.4; DB 9; Length 10662;
Best Local Similarity 53.2%; Pred. No. 8.4e-29;
Matches 433; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGCCAGATCATCCCTACCAACCCCAAGGCAACACAGCAGT 60
DB 8961 CAGCGGGGGAACCTTCGTGGCCAAAGATCGTCCCTATGTGCGGAGGGCAAGCGGGGT 9020
QY 61 GCTGCGCGAATACAGGCGCCCTCAAGGGCTGGCGCCACCGCCACCTGGCCAGCTGCACGC 120
DB 9021 CTTGAGGAGTACAGAGTGTGGGACCCCTGCACACAGCGGATCATGTCCCTGCACGA 9080
QY 121 AGCCTACCTAGCCCGGGCAGCTGTGCTCATCTTGGAGCTGTGCTCTGGGCGCGAGCT 180
DB 9081 GGCTATACATCACCCCTCGGTACCTCGTGTCTATTGTGTAGAGCTGTGGCAACCGGGAAT 9140
QY 181 GCTCCCTCGCTGGCGAGAGGGCCCTCCTACTCAGAAATCCGAGGTGAAGGACTACTGTG 240
DB 9141 CCTCTGTGGCTCAGTGACAGGTTCGGGTATCTTGAGGATGACGTGGCCACTTACATGCT 9200

QY 241 GCAGATGTTAGTGCACCCAGTACTCTGCAAAACGAGACATCTCTGCACCTGACCTGAG 300
DB 9201 GCAGCTGCTACAAGGCTGGACTACTCTCCACGGCCACCACGTCCTCCACCTAGATCAA 9260
QY 301 GTCCGAGACATGATCATCACCGAATAACAACCTGCTCAAGGTCGTGGACCTGGCAATGC 360
DB 9261 GCCAGACAACCTGCTGCTGGCCCTCGACAAATGCCCTCAAGATTGGGACTTTGGCAGTGC 9320
QY 361 ACAGAGCTTCAGCCAGGAGAAGTGTGCTCCCTCAGACAAAGTTCAAGGACTACTAGAGAC 420
DB 9321 CCAGCCCTACAACCCCGAGGCTTAGGCCCTTGCCCAACCGCAGGGGACGCTGGAGTT 9380
QY 421 CATGGCTCCAGAGCTCCTCGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480
DB 9381 CATGGCTCCGAGATGGTGAAGGGAGAACCCATTCGGCTCTGCCACGACATCTGGGGAGC 9440
QY 481 CGGTGTGACAGCCTTCATCATGTGAGCGCCGAGTACCCCGTGTAGCAGCGAGGGTGACG 540
DB 9441 GGGTGTGCTCACTTACATATGCTCAGTGGAGCTCCCGCTTCTATGAGCCAGACCCCA 9500
QY 541 CGACCTGCGAGAGGACTGGGCAAGGGGTGGTCCGGCTGAGCGCGTGTCTAGCGGGCT 600
DB 9501 GGAACCGGAGGCTCGGATTGTGGGGGGCGCTTTGATGCTTCCAGCTGTACCCCAATAC 9560
QY 601 GTCCGGGGGGCGCGTGGCCTTCTGTGCGCAGCACTCTGTGCGCCACGCTTGGGGCGCGCC 660
DB 9561 ATCCGAGCGCCACCTCTTCTTGGGAAAGTTCTCTCTGTATACATCCCTGGAGCGGCG 9620
QY 661 CTCGGCTCAGCTGCTGCGAGTCCCGCTGTGCTGTGAGAGAGGGCGCGCTGTTCGCG 720
DB 9621 CTCCTGCGAGGACTGCTGCGCCACCCATGGTTGCGAGGACGCTTACTGATGAAGTGGC 9680
QY 721 GCCCGGGCGCTGACCTTCCCTACCGGGGGTGGCGCTTTCGTGCGCAATCGCGAGAA 780
DB 9681 CCGCCAGACCTCACCTTACACCAACCGGCTCAAGGAGTTCTCTGGGGGAGCAGCGCG 9740
QY 781 GAGACGCGCGCTGCTGTATCAAGAGGCAACACCTG 814
DB 9741 GCGCGGGCTGAGGCTGCCACCGCCACCAAGGTG 9774

Search completed: September 19, 2004, 15:42:26
Job time : 404.361 secs

This Page Blank (uspto)

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 12:53:28 ; Search time 158.742 Seconds
(without alignments)
10303.209 Million cell updates/sec

Title: US-10-077-130-4_COPY_16862_17246

Perfect score: 385
Sequence: 1 ccgaggagagatctttgaca.....acctgcagcaccctggagcgc 385

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_29Jan04:*
- 1: Geneseqn1980s:*
 - 2: Geneseqn1990s:*
 - 3: Geneseqn2000s:*
 - 4: Geneseqn2001as:*
 - 5: Geneseqn2001bs:*
 - 6: Geneseqn2002s:*
 - 7: Geneseqn2003as:*
 - 8: Geneseqn2003bs:*
 - 9: Geneseqn2003cs:*
 - 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 385 | 100.0 | 7928 | 5 | AAC62287 cDNA enco |
| 2 | 385 | 100.0 | 8106 | 7 | ABX11641 Human ser |
| 3 | 385 | 100.0 | 24120 | 7 | ABX11642 Human ser |
| 4 | 377 | 97.9 | 628 | 3 | AAC74767 Human ORF |
| 5 | 377 | 97.9 | 628 | 6 | ABN75689 Human ORF |
| 6 | 42.6 | 11.1 | 4180 | 6 | ABK84118 Human CDN |
| 7 | 42.6 | 11.1 | 4282 | 6 | ABX51830 Novel hum |
| 8 | 42.6 | 11.1 | 4366 | 7 | ABX63668 Human cDN |
| 9 | 41.4 | 10.8 | 1803 | 5 | AAC84312 Human EXC |
| 10 | 39.8 | 10.3 | 1278 | 7 | ACA23561 Prokaryot |
| 11 | 39.4 | 10.2 | 455 | 8 | ACH34588 Human end |
| 12 | 39.4 | 10.2 | 474 | 8 | ACH34675 Human end |
| 13 | 39.4 | 10.2 | 1302 | 6 | ABS55941 DNA topoi |
| 14 | 39.4 | 10.2 | 1354 | 4 | AAS26907 Human cDN |
| 15 | 39.4 | 10.2 | 1354 | 6 | ABQ54643 Human ova |
| 16 | 39.4 | 10.2 | 1371 | 4 | AAS26834 Human CDN |
| 17 | 39.4 | 10.2 | 1720 | 7 | ABX05065 Human nov |
| 18 | 39.4 | 10.2 | 2040 | 9 | ADB63018 Human cDN |
| 19 | 39.4 | 10.2 | 2151 | 5 | ABAI5210 Human ner |
| 20 | 39.4 | 10.2 | 2156 | 5 | ABAI5209 Human ner |
| 21 | 39.4 | 10.2 | 125401 | 4 | AAI17186 Streptomy |
| 22 | 38.8 | 10.1 | 1131 | 7 | ACA25328 Prokaryot |
| 23 | 38.6 | 10.0 | 307 | 6 | ABL79596 Human Ova |

| | | | | | | |
|----|------|------|-------|---|----------|--------------------|
| 24 | 38.6 | 10.0 | 21185 | 3 | AAA63350 | Aaa63350 Streptomy |
| 25 | 38.6 | 10.0 | 63164 | 3 | AAA63348 | Aaa63348 Streptomy |
| 26 | 38 | 9.9 | 1032 | 7 | ADA69587 | Ada69587 Rice gene |
| 27 | 37.8 | 9.8 | 5154 | 5 | AAS84859 | Aas84859 DNA enco |
| 28 | 37.8 | 9.8 | 5751 | 4 | AAK51986 | Aak51986 Human pol |
| 29 | 37.6 | 9.8 | 3525 | 4 | ABL09775 | Ab109775 Drosophil |
| 30 | 37.6 | 9.8 | 8848 | 4 | ABL09774 | Ab109774 Drosophil |
| 31 | 37.2 | 9.7 | 1200 | 3 | AAA30024 | Aaa30024 Human PRO |
| 32 | 37.2 | 9.7 | 1200 | 3 | AAA77525 | Aaa77525 Human PRO |
| 33 | 37.2 | 9.7 | 1200 | 3 | ADC78342 | Adc78342 Human PRO |
| 34 | 37.2 | 9.7 | 1200 | 4 | AAF72376 | Aaf72376 Human PRO |
| 35 | 37.2 | 9.7 | 1200 | 4 | AAC87033 | Aac87033 Nucleotid |
| 36 | 37.2 | 9.7 | 1200 | 4 | AAS21380 | Aas21380 Human CDN |
| 37 | 37.2 | 9.7 | 1200 | 4 | AAF58494 | Aaf58494 PRO187 co |
| 38 | 37.2 | 9.7 | 1200 | 7 | ACA58913 | Aca58913 Human PRO |
| 39 | 37.2 | 9.7 | 1200 | 7 | ACD23989 | Acd23989 Novel hum |
| 40 | 37.2 | 9.7 | 1200 | 7 | ACA58310 | Aca58310 cDNA enco |
| 41 | 37.2 | 9.7 | 1200 | 7 | ACD42383 | Novel hum |
| 42 | 37.2 | 9.7 | 1200 | 7 | ACA67130 | Novel hum |
| 43 | 37.2 | 9.7 | 1200 | 7 | ACA60017 | Novel hum |
| 44 | 37.2 | 9.7 | 1200 | 7 | ACD07417 | Novel hum |
| 45 | 37.2 | 9.7 | 1200 | 7 | ACA03739 | Novel hum |

ALIGNMENTS

RESULT 1
AAC62287

ID AAC62287 standard; cDNA; 7928 BP.

XX AAC62287;

DT 19-MAR-2001 (first entry)

XX cDNA encoding a splice variant of a signal transduction polypeptide.

XX Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
XX congestive heart failure; dilated congestive cardiomyopathy;
XX hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
XX mitral valve disease; aortic valve disease; tricuspid valve disease;
XX myocardial infarction; cardiac arrhythmia; arteriosclerosis;
XX atherosclerosis; cardiac tumour; microbial infection; splice variant; ss.
XX Homo sapiens.

Key Location/Qualifiers
CDS 60..7850
/tag= a
/product= "signal transduction polypeptide H19G5 splice variant"

FT FT WO200063381-A1.

PD 26-OCT-2000.

PF 11-APR-2000; 2000WO-US009488.

PR 16-APR-1999; 99US-0129553P.

XX (SCIO-) SCIOS INC.

PA Zeng W, Stanton L, Kong H;

PI WPI; 2001-007013/01.

XX P-PSDB; AAB30569.

PT Novel h19G5 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
XX cardiac diseases, and additional mediators of signal transduction.

PS Claim 4; Page 65-68; 81pp; English.

| | |
|----|--|
| CC | The present sequence encodes a splice variant of human in signal |
| CC | transduction polypeptide. The polypeptide is designated H19G5. The |
| CC | protein is capable of regulating signal transduction and exhibits kinase |
| CC | activity. The H19G5 transcript is expressed in the heart. H19G5 |
| CC | polypeptides and polynucleotides are useful for preventing or treating a |
| CC | cardiac disease, such as congestive heart failure, dilated congestive |
| CC | cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, |
| CC | mitral valve disease, aortic valve disease or tricuspid valve disease, |
| CC | angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, |
| CC | arterial or renovascular hypertension, arteriosclerosis, atherosclerosis |
| CC | and cardiac tumours in humans. The polypeptide is also useful for |
| CC | detecting the expression of a protein capable of regulating signal |
| CC | transduction or the expression of a protein capable of acting as a donor |
| CC | or acceptor molecule of a phosphate group. The monoclonal antibodies can |
| CC | be used as probes for detecting discrete antigens expressed by tissue or |
| CC | cell samples, and therefore used in humans for localization and |
| CC | monitoring of microbial infection |
| XX | |
| SQ | Sequence 7928 BP; 1535 A; 2644 C; 2488 G; 1261 T; 0 U; 0 Other; |
| | Query Match 100.0%; Score 385; DB 5; Length 7928; |
| | Best Local Similarity 100.0%; Pred. No. 7.2e-87; |
| | Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| QY | 1 CGAGGGCAGATCTTTGACATCTACGTGTTCACCGTGTACTACCTGCCCTTAGGGGCTGA 60 |
| DB | 734 CGAGGCAGAGATCTTTGACATCTACGTGTTCACCGTGTACTACCTGCCCTTAGGGGCTGA 793 |
| QY | 61 GCAGAGTGCCATCACGCTCGGGGAGGCCAGTAGTATGTGGAGGTCTCTGGATGCAGGCCACC 120 |
| DB | 794 GCAGAGTGCCATCACGCTCGGGAGGCCAGTAGTATGTGGAGGTCTCTGGATGCAGGCCACC 853 |
| QY | 121 ACTCGCTGGGTTGTTCGCACAAGAGCCCAAGTAGTCAGCCCTCACGGCAGGGCTGGGT 180 |
| DB | 854 ACTCGCTGGGTTGTTCGCACAAGAGCCCAAGTAGTCAGCCCTCACGGCAGGGCTGGGT 913 |
| QY | 181 GTCCACAGCTTACCTGGACAGAGGCTCAAAGCTGTCACTGAGTGGGGGGCCGCTGAGGC 240 |
| DB | 914 GTCCACAGCTTACCTGGACAGAGGCTCAAAGCTGTCACTGAGTGGGGGGCCGCTGAGGC 973 |
| QY | 241 CCTGAGTTCCCTGGGGAGGCTGTGTCTGAAGACGAATACAAAGCAGAAGCTGAGCTCTGT 300 |
| DB | 974 CCTGAGTTCCCTGGGGAGGCTGTGTCTGAAGACGAATACAAAGCAGAAGCTGAGCTCTGT 1033 |
| QY | 301 GATCCAGGAGCTGTGAGTTCTGAGCAGCGCTTCGTGGAGGAGCTGCAGTTCCTGCAGAG 360 |
| DB | 1034 GATCCAGGAGCTGTGAGTTCTGAGCAGCGCTTCGTGGAGGAGCTGCAGTTCCTGCAGAG 1093 |
| QY | 361 CCACCACTTCGAGCACTTGGAGCGC 385 |
| DB | 1094 CCACCACTTCGAGCACTTGGAGCGC 1118 |

RESULT 2

RESUL 2
ABX11641

ABX11641
ID ABX11641 standard; cDNA: 8106 BP.

XX
TO THE SECRETARY OF THE ARMY, WASHINGTON, D.C.

AC ABX11641:

XX
XX
XXXXXXXXXXXX

DT 09-MAY-2003 (first entry)

DI 03-PMI-2003 (11130 entry)
XX

Human serine/threonine or protein

DE
HUMAN
SERUM
/ CHLOROFORM
OR PROTEIN
XX

Human: ss: gene: serine/threonine

KW Humall; ss; gene; serine/ chalcone
KW cardiovascular disease: heart fai

KW cardiovascular disease; heart failure
KW blood vessel disorder; atherosclerosis

KW blood vessel disorder; atherosclerosis
KW blood platelet disorder; thrombocytopenia

KW blood platelet disorder; thrombocytopenic anaemia: cellular proliferation

KW haemolytic anaemia; cellular protein kinase disorder; autoimmune

KW protein kinase disorder; autoimmune psoriasis; inflammatory bowel disease

multiple sclerosis
psoriasis; inflammatory bowel dis

KW multiple sclerosis.
XX

11

| | | |
|----|--------------------------------------|---|
| XX | Key | Location/Qualifiers |
| FH | 5'UTR | 1..71 |
| FT | | /*tag= a |
| FT | | 72..7964 |
| FT | CDS | /*tag= b |
| FT | | /product= "Kinase 59079" |
| FT | | /note= "This CDS is specifically claimed in claim 2" |
| FT | 3'UTR | 7965..8106 |
| FT | | /*tag= c |
| XX | | US2002168742-A1. |
| XX | | 14-NOV-2002. |
| XX | | 15-FEB-2002; 2002US-00077130. |
| XX | | 15-FEB-2001; 2001US-0269201P. |
| XX | | (MILL-) MILLENNIUM PHARM INC. |
| PA | | Kapeller-Libermann R, Acton SL; |
| PI | | WPI; 2003-298729/29. |
| DR | | P-PSDS; ABG76186. |
| DR | | |
| XX | | Novel isolated human protein kinase, designated 59079 or 12599 |
| PT | | polypeptide, useful as diagnostic and therapeutic agents for preventing |
| PT | | cardiovascular diseases, proliferative disorders, and protein kinase |
| PT | | disorders. |
| XX | | |
| XX | Claim 2; Page 39-48; 119pp; English. | |
| XX | | |
| CC | | The invention relates to an isolated human serine/threonine or protein |
| CC | | kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule |
| CC | | comprising at least 85% identity to the nucleic acids appearing as |
| CC | | ABX11641 and ABX1642 or their complement, a naturally occurring variant |
| CC | | of the kinases or their fragments. Also included are a non-human host |
| CC | | cell containing the nucleic acids, an antibody specific for the proteins, |
| CC | | identifying a compound which binds to the kinase (by contacting the |
| CC | | kinase or a cell expressing the kinase with a test compound and |
| CC | | determining whether the kinase binds to the test compound) and modulating |
| CC | | the activity of kinase using the identified compound. The kinases and |
| CC | | their encoding nucleic acids are useful as diagnostic and therapeutic |
| CC | | agents for preventing a disease or condition associated with an aberrant |
| CC | | or unwanted 59079 or 12599 activity in a subject, including |
| CC | | cardiovascular diseases such as heart failure, and myocardial infarction; |
| CC | | disorders involving blood vessels such as atherosclerosis, and Kaposi's |
| CC | | sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia, |
| CC | | Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders |
| CC | | such as cancer; and protein kinase disorders such as autoimmune |
| CC | | disorders, diabetes mellitus, psoriasis, inflammatory bowel disease, |
| CC | | rheumatoid arthritis, and multiple sclerosis (many examples of diseases |
| CC | | and disorders are included in the specification). The kinases, their |
| CC | | encoding nucleic acids and antibodies are useful in screening assays, |
| CC | | detection assays (e.g. forensic biology), and predictive medicine (e.g. |
| CC | | diagnostic assays, prognostic assays, and monitoring clinical trials and |
| CC | | pharmacogenomics). The kinases and their encoding nucleic acids are |
| CC | | useful as query sequences to perform a search against public databases to |
| CC | | identify other family members or related sequences. The present sequence |
| CC | | encodes the kinase 59079 |

XX
SQ Sequence 8106 BP: 1603 A: 2684 C: 2530 G: 1289 T: 0 U: 0 Other:;

Human serine/threonine or protein kinase 59079, cDNA.

Human: ss: gene: serine/threonine kinase: protein kinase: 59079; KW

KW Humall; ss; gene; serine/threonine kinase; protein kinase; cardiovascular disease; heart failure; myocardial infarction;

KW blood vessel disorder: atherosclerosis: Kaposi's sarcoma: KW cardiovascular disease; heart failure; myocardial infarction; KW

blood vessel disorder; atherosclerosis; kaposis sarcoma;
blood platelet disorder: thrombocytopaenia: leukaemia: Hodgkin's

KW blood platelet disorder; thrombocytopenia; leukaemia; myeloid
KW haemolytic anaemia: cellular proliferative disorder: cancer:


protein kinase disorder: autoimmune disorder: diabetes mellitus: haemolytic anaemia; cellular proliferative disorder; cancer; kw

protein kinase disorder; autoimmune disorder; diabetes mellitus; psoriasis: inflammatory bowel disease: rheumatoid arthritis:

psoriasis; inflammatory bowel disease; rheumatoid arthritis;
multiple sclerosis

multiple sclerosis.
KW
XX























Db 908 GCAGGATGCCATCACGCTGGGGAGGCCAGTATGTGGAGTCTGTGGATGAGCCACCC 967
QY 121 ACTGCGCTGGCTGTGTCGCGACCAAGCCCAAGTCCAGCCCTCACGCGAGGGCTGGT 180
Db 968 ACTGCGCTGGCTGTGTCGCGACCAAGCCCAAGTCCAGCCCTCACGCGAGGGCTGGT 1027
QY 181 GTCACAGCCTACCTGGAAGAGGCTCAAGTCTGACCTGAGTGGGGGGCGCTGAGGC 240
Db 1028 GTCACAGCCTACCTGGAAGAGGCTCAAGTCTGACCTGAGTGGGGGGCGCTGAGGC 1087
QY 241 CCCTGAGTTCCTGGGAGGCTGTGTCGAGAGCAATACAGGCAAGGCTGACTCTGT 300
Db 1088 CCCTGAGTTCCTGGGAGGCTGTGTCGAGAGCAATACAGGCAAGGCTGACTCTGT 1147
QY 301 GATCCAGAGCTGTGAGTCTTCTGAGCAGGCCCTTCTGGAGGAGCTGCAGTTCTTCGAGAG 360
Db 1148 GATCCAGAGCTGTGAGTCTTCTGAGCAGGCCCTTCTGGAGGAGCTGCAGTTCTTCGAGAG 1207
QY 361 CCACCACTGCAGCAGCTGGAGCGC 385
Db 1208 CCACCACTGCAGCAGCTGGAGCGC 1232

RESULT 3
ID ABX11642 standard; cDNA; 24120 BP.
XX ABX11642;
AC ABX11642;
DT 09-MAY-2003 (first entry)
XX Human serine/threonine or protein kinase 12599, cDNA.
DE Human; ss; gene; serine/threonine kinase; protein kinase; 12599;
KW cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;
KW haemolytic anaemia; cellular proliferative disorder; cancer;
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.
XX Homo sapiens.
XX Key
XX 5'UTR 1..71 Location/Qualifiers
FT /tag= a
FT 72..23978
FT /tag= b
FT /product= "Kinase 12599"
FT /note= "This CDS is specifically claimed in claim 2"
FT 23979..24120
FT /tag= c
XX 3'UTR
XX US2002168742-A1.
XX 14-NOV-2002.
XX 15-FEB-2002; 2002US-00077130.
XX 15-FEB-2001; 2001US-0269201P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Kapeller-Libermann R, Acton SL;
XX WPI; 2003-298729/29.
XX P-FSDB; ABG76187.
XX Novel isolated human protein kinase, designated 59079 or 12599
XX polypeptide, useful as diagnostic and therapeutic agents for preventing
XX cardiovascular diseases, proliferative disorders, and protein kinase

disorders.
XX Claim 2; Page 58-84; 119pp; English.
XX The invention relates to an isolated human serine/threonine or protein
XX kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule
XX comprising at least 85% identity to the nucleic acids appearing as
XX ABX11641 and ABX11642 or their complement, a naturally occurring variant
XX of the kinases or their fragments. Also included are a non-human host
XX cell containing the nucleic acids, an antibody specific for the proteins,
XX identifying a compound which binds to the kinase (by contacting the
XX kinase or a cell expressing the kinase with a test compound and
XX determining whether the kinase binds to the test compound) and modulating
XX the activity of kinase using the identified compound. The kinases and
XX their encoding nucleic acids are useful as diagnostic and therapeutic
XX agents for preventing a disease or condition associated with an aberrant
XX or unwanted 59079 or 12599 activity in a subject, including
XX cardiovascular diseases such as heart failure, and myocardial infarction;
XX disorders involving blood vessels such as atherosclerosis, and Kaposi's
XX sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia,
XX Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders
XX such as cancer; and protein kinase disorders such as autoimmune
XX disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,
XX rheumatoid arthritis, and multiple sclerosis (many examples of diseases
XX and disorders are included in the specification). The kinases, their
XX encoding nucleic acids and antibodies are useful in screening assays,
XX detection assays (e.g. forensic biology), and predictive medicine (e.g.
XX diagnostic assays, prognostic assays, and monitoring clinical trials and
XX pharmacogenomics). The kinases and their encoding nucleic acids are
XX useful as query sequences to perform a search against public databases to
XX identify other family members or related sequences. The present sequence
XX encodes the kinase 12599
XX
SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;
Query Match 100.0%; Score 385; DB 7; Length 24120;
Best Local Similarity 100.0%; Pred. No. 9.2e-87;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGAGCGAGATCTTGACATCTACGTGGTCCACCGCTACTCTGCTGGGGCTGA 16921
Db 16862 CCGAGCGAGATCTTGACATCTACGTGGTCCACCGCTACTCTGCTGGGGCTGA 16921
QY 61 GCAGGATGCCATCACGCTCGGGAGGCCAGTATGTGGAGGTCCTGGATGAGCCACCC 120
Db 16922 GCAGGATGCCATCACGCTCGGGAGGCCAGTATGTGGAGGTCCTGGATGAGCCACCC 16981
QY 121 ACTGCGCTGGCTGTGTCGCGACCAAGCCCAAGTCCAGCCCTCACGCGAGGGTGGT 180
Db 16982 ACTGCGCTGGCTGTGTCGCGACCAAGCCCAAGTCCAGCCCTCACGCGAGGGTGGT 17041
QY 181 GTCACAGCCTACCTGGAAGAGGCTCAAGTCTGACCTGAGTGGGGGGCGCTGAGGC 240
Db 17042 GTCACAGCCTACCTGGAAGAGGCTCAAGTCTGACCTGAGTGGGGGGCGCTGAGGC 17101
QY 241 CCCTGAGTTCCTGGGAGGCTGTGTCGAGAGCAATACAGGCAAGGCTGAGCTCTGT 300
Db 17102 CCCTGAGTTCCTGGGAGGCTGTGTCGAGAGCAATACAGGCAAGGCTGAGCTCTGT 17161
QY 301 GATCCAGAGCTGTGAGTCTTCTGAGCAGGCCCTTCTGGAGGAGCTGCAGTTCTTCGAGAG 360
Db 17162 GATCCAGAGCTGTGAGTCTTCTGAGCAGGCCCTTCTGGAGGAGCTGCAGTTCTTCGAGAG 17221
QY 361 CCACCACTGCAGCAGCTGGAGCGC 385
Db 17222 CCACCACTGCAGCAGCTGGAGCGC 17246

RESULT 4
AAC74767
ID AAC74767 standard; cDNA; 628 BP.
XX AAC74767;
AC AAC74767;

| | | |
|----|--|---------------|
| XX | 08-FEB-2001 | (first entry) |
| DT | Human ORFX ORE322 polynucleotide sequence SEQ ID NO:643. | |
| XX | | |
| DE | Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; | |
| XX | vulnary; antiposariatic; antiparkinsonian; neutropic; neuroprotective; | |
| KW | anticonvulsant; osteopathic; antiarthritic; coagulant; vasotropic; cardiant; | |
| KW | immunostimulant; thrombolytic; immunosuppressive; antidiabetic; | |
| KW | hypotensive; dermatological; immunosuppressive; antihypertensive; | |
| KW | antiviral; antibacterial; antifungal; antirheumatic; antithyroid; | |
| KW | antianemic; gene therapy; cancer; proliferative disorder; hypertension; | |
| KW | neurodegenerative disorder; osteoarthritis; graft vs host disease; | |
| KW | cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; | |
| KW | cholesterol ester storage; systemic lupus erythematosus; infection; | |
| KW | severe combined immunodeficiency; malaria; autoimmune disorder; asthma; | |
| KW | allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; | |
| KW | bone damage; cartilage damage; antiinflammatory disease; coagulation; | |
| KW | thrombosis; contraceptive; ss. | |
| OS | Homo sapiens. | |
| XX | | |
| XX | WO200058473-A2. | |
| PN | | |
| XX | | |
| PD | 05-OCT-2000. | |
| XX | | |
| PF | 31-MAR-2000; 2000WO-US008621. | |
| XX | | |
| PR | 31-MAR-1999; 99US-0127607P. | |
| PR | 02-APR-1999; 99US-0127636P. | |
| PR | 05-APR-1999; 99US-0127728P. | |
| PR | 30-MAR-2000; 2000US-00540763. | |
| XX | | |
| PA | (CURA-) CURAGEN CORP. | |
| PI | Shimkets RA, Leach M; | |
| PI | | |
| PT | WPI: 2000-602362/57. | |
| DR | P-PSDB; AAB40358. | |
| XX | | |
| XX | Novel nucleic acids and peptides derived from open reading frame X, | |
| PT | useful for treating e.g. cancers, proliferative disorders, | |
| PT | neurodegenerative disorders and cardiovascular disease. | |
| XX | | |
| PS | Claim 5; Page 764; 5507pp; English. | |
| XX | | |
| CC | AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, | |
| CC | which represent the human ORFX open reading frames 1 to 3161. The ORFX | |
| CC | sequences have activities such as: cytostatic; hepatotropic; vulnary; | |
| CC | antiposariatic; antiparkinsonian; neutropic; neuroprotective; osteopathic; | |
| CC | anticonvulsant; antiarthritic; immunosuppressive; immunostimulant; | |
| CC | cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; | |
| CC | dermatological; immunosuppressive; antiinflammatory; antibacterial; | |
| CC | antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The | |
| CC | sequences can be used for determining the presence of or predisposition | |
| CC | to, or preventing or treating pathological conditions associated with an | |
| CC | ORFX-associated disorder. The nucleic acids can be used to express ORFX | |
| CC | proteins in gene therapy vectors. The proteins and nucleic acids may be | |
| CC | used to treat cancers, proliferative disorders, neurodegenerative | |
| CC | disorders, osteoarthritis, graft vs host disease, cardiovascular disease, | |
| CC | diabetes mellitus, hypertension, hypothyroidism, cholesterol ester | |
| CC | storage, systemic lupus erythematosus, severe combined immunodeficiency | |
| CC | (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune | |
| CC | disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and | |
| CC | cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to | |
| CC | enhance coagulation; to inhibit thrombosis; and as a contraceptive | |
| XX | | |
| XX | Sequence 628 BP; 122 A; 189 C; 200 G; 116 T; 0 U; 1 Other; | |

| | | | |
|----------|--|--|-----|
| QY | 9 | AGATCTTTGACATCTACGTGTGTACCGCTGACTACCTGCGCCCTAGGGCTGAGCAGGATG | 68 |
| Db | 2 | AGATCTTTGACATCTACGTGTGTGTACCGCTGACTACCTGCGCCCTAGGGCTGAGCAGGATG | 61 |
| QY | 69 | CCATCACGCTGCGGGAAGCCAGTATGTGGAGGTCTCTGGATGACGCCACCCACTGCGCT | 128 |
| Db | 62 | CCATCACGCTGCGGGAAGGCCAGTATGTGGAGGTCTCTGGATGACGCCACCCACTGCGCT | 121 |
| QY | 129 | GGCTGTGTGGGCACCAAGGCCCAAGTCCAGGCCCTCACGGCAGGGCTGGGTGTACACAG | 188 |
| Db | 122 | GGCTGTGTGGGCACCAAGGCCCAAGTCCAGGCCCTCACGGCAGGGCTGGGTGTACACAG | 181 |
| QY | 189 | CCTACTCTGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGGCGCTGAGGCCCTTGACT | 248 |
| Db | 182 | CCTACTCTGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGGCGCTGAGGCCCTTGACT | 241 |
| QY | 249 | TCCTGGGGAGGCTGTGTCTGAAGCAAGATACAAAGCAGAGCTGAGCTCTGTGATCCAG | 308 |
| Db | 242 | TCCTGGGGAGGCTGTGTCTGAAGCAAGATACAAAGCAGAGCTGAGCTCTGTGATCCAG | 301 |
| QY | 309 | AGCTGCTGAGTTCTGAGCAGGCGCTTCGTGGAGGAGCTGCGAGTTCCTGCAGAGCCACCC | 368 |
| Db | 302 | AGCTGCTGAGTTCTGAGCAGGCGCTTCGTGGAGGAGCTGCGAGTTCCTGCAGAGCCACCC | 361 |
| QY | 369 | TGCAGCACCTTGGAGCGC | 385 |
| Db | 362 | TGCAGCACCTTGGAGCGC | 378 |
| RESULT 5 | | | |
| ABN75689 | | | |
| ID | ABN75689 standard; cDNA; 628 BP. | | |
| XX | ABN75689; | | |
| AC | 08-JUL-2002 (first entry) | | |
| DT | Human ORF636 cDNA, SEQ ID NO:1271. | | |
| DE | Human; ORF; open reading frame; ORFX; drug screening; diagnosis; | | |
| XX | disease monitoring; cytokine; cell proliferation; cell differentiation; | | |
| KW | immune modulation; haematopoiesis regulation; tissue growth; | | |
| KW | angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic; | | |
| KW | thrombolytic; tumour inhibition; bodily characteristic; fertility; | | |
| KW | behaviour; cancer; proliferative disorder; neurological disorder; | | |
| KW | cardiovascular disease; immune system disorder; organ transplantation; | | |
| KW | tissue growth disorder; tissue regeneration disorder; diabetes mellitus; | | |
| KW | hypochloridism; cholesterol ester storage disease; infection; vulnuary; | | |
| KW | vasorropic; antiparietal; antidiabetic; cyostatic; neutropic; | | |
| KW | neuroprotective; antithaerosteric; anticoagulant; thrombolytic; | | |
| KW | cardiant; hypotensive; antithyroid; antinflammatory; immunomodulator; | | |
| KW | dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss. | | |
| OS | Homo sapiens. | | |
| XX | WO200190366-A2. | | |
| FN | 29-NOV-2001. | | |
| XX | PD | | |

XX Novel human polypeptides and polynucleotides useful for diagnosing, preventing and treating cardiovascular disease, neurodegenerative, hypertroliferative disorders and disorders related to organ

Query Match 97.9%; Score 377; DB 3; Length 628;
Best Local Similarity 100.0%; Pred. No. 4.2e-85;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0

CC expressed in granulocytes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 4180 BP; 787 A; 1249 C; 1285 G; 859 T; 0 U; 0 Other;

Query Match 11.1%; Score 42.6; DB 6; Length 4180;
 Best Local Similarity 48.5%; Pred. No. 0.67;
 Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGTGTCCACAGCTACTGGACAGGA 203
 |||||
 Db 313 AGCCCGAGGACGACGCCGGTCCAGGACGGGCTGGATGGCCGCCACACAGTGGAGG 372

QY 204 GGCTCAAGCTGCACCTAGTGGGGGCGGCTGAGGCCCTGAGTTCCTGGGAGGCTG 263
 |||||
 Db 373 AGCTGAAGCGCGCTGTGAGCGCGGCGAGCTGGAGGCGGCGCGCTGTGGCGTGG 432

QY 264 TGTCTGAAGACGAATACAAAGCAAGGCTGAGCTCTGTGATCCAGGAGCTGTGAGTTC 323
 |||||
 Db 433 AGCGGAGCTGGCGGCGCGGCGGCGGCGGCGGCTGTGAGGAGGAGCTGTGGCGGC 492

QY 324 AGCAGGCTTGTGGAGAGCTGAGTTCCTTGGAGAGCCACCTGACGACCTGGAGC 383
 |||||
 Db 493 GCCAGAGCAAGTGGAGGCGCTGTACGAGCTGCTGCGGACCAAGGTGCTGGCGGTGCTG 552

QY 384 G 384
 553 G 553

RESULT 7

ABSS1830

ID ABSS1830 standard; cDNA; 4282 BP.

XX AC ABSS1830;

XX DT 05-NOV-2002 (first entry)

XX DE Novel human thrombopoietin splice variant cDNA, NV-16.

XX KW Human; splice variant; chromosome identification; cancer; thrombopoietin;

XX KW thrombocytopenia; transporter protein; gene; ss.

XX OS Homo sapiens.

XX PN US2002068342-A1.

XX PD 06-JUN-2002.

XX PF 08-FEB-2001; 2001US-00778927.

XX PR 09-FEB-2000; 2000IL-00134453.

XX PR 29-MAR-2000; 2000IL-00135341.

XX PA (KHOS/) KHOSRAVI R.

XX PA (BERN/) BERNSTEIN J.

XX PI Khosravi R, Bernstein J;

XX DR WPI; 2002-582479/62.

XX DR P-PSDB; ABG70357.

XX PT Novel nucleic acid sequences that code for variants obtained by

XX PT alternative splicing; homologs of known thrombopoietins, and homologs of

XX PT transporter proteins, for treating diseases or disorders e.g.

XX PT thrombocytopenia.

XX PS Claim 1; Fig 16; 114pp; English.

XX CC The invention relates to novel nucleic acid (I) sequence of an

XX CC alternative splice variant and the corresponding cDNA sequence, all recorded by

CC (I). (I) is useful for a variety of diagnostic purposes and to detect and
 CC quantitate expression of the sequences in patient's cells e.g. biopsied
 CC tissues, for chromosome identification and for therapeutic purposes. The
 CC novel variants also serve for detection i.e. their presence or level
 CC indicates disease, disorder, pathological or normal condition or
 CC alternatively the ratio between the level of variants and the level of
 CC original sequence from which they were varied or the ratio to other
 CC variants may be indicative of a disease, disorder or pathological
 CC condition. The variants are useful for the development of pharmaceuticals
 CC for various pathological conditions in which cell cycle is not normal,
 CC notably cancer. Thrombopoietin (TPO) product is useful in
 CC thrombocytopenia, transporter protein (TH) product is useful in
 CC conjunction with imaging substances for detection and imaging purposes.
 CC The variant product, the TPO product or the TH product, its catalytic or
 CC immunogenic fragments or oligopeptides are useful for screening
 CC therapeutic compounds in a variety of drug screening techniques. ABSS1815
 CC .ABSS1855 represent novel human splice variant coding sequences of the
 CC invention

XX SQ Sequence 4282 BP; 816 A; 1279 C; 1303 G; 883 T; 0 U; 1 Other;

Query Match 11.1%; Score 42.6; DB 6; Length 4282;
 Best Local Similarity 48.5%; Pred. No. 0.67;
 Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGTGTCCACAGCTACTGGACAGGA 203
 |||||
 Db 526 AGCCCGAGGACGACGCCGGTCCAGGACGGGCTGGATGGCCGCCACACAGTGGAGG 585

QY 204 GGCTCAAGCTGCACCTAGTGGGGGCGGCTGAGGCCCTGAGTTCCTGGGAGGCTG 263
 |||||
 Db 586 AGCTGAAGCGCGCTGTGAGCGCGGCGAGCTGGAGGCGGCGGCGCTGTGGCGCTGG 645

QY 264 TGTCTGAAGACGAATACAAAGCAAGGCTGAGCTCTGTGATCCAGGAGCTGTGAGTTC 323
 |||||
 Db 646 AGCGGAGCTGGCGGCGGCGGCGGCGGCGGCTGTGAGGAGGAGCTGTGGTGGCGC 705

QY 324 AGCAGGCTTGTGGAGAGCTGAGTTCCTTGGAGAGCCACCTGACGACCTGGAGC 383
 |||||
 Db 706 GCCAGAGCAAGTGGAGGCGCTGTACGAGCTGCTGCGGACCAAGGTGCTGGCGGTGCTG 765

QY 384 G 384
 766 G 766

RESULT 8

ABX63668

ID ABX63668 standard; cDNA; 4366 BP.

XX AC ABX63668;

XX DT 26-FEB-2003 (first entry)

XX DE Human cDNA #668 differentially expressed in activated vascular tissue.

XX KW Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;

XX KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;

XX KW gene therapy; vascular disease; cancer; coronary; artery disease;

XX KW hypertension; diabetes; pre-eclampsia; restenosis;

XX KW ischaemia-reperfusion injury; stroke.

XX OS Homo sapiens.

XX PN US2002137081-A1.

XX PD 26-SEP-2002.

XX PF 08-JAN-2002; 2002US-00044090.

XX PR 28-JUL-2000; 2000US-0222469P.

XX PR 08-JAN-2001; 2001US-0260483P.

```

Query Match      11.1%; Score 42.6; DB 7; Length 4366;
Best Local Similarity 48.5%; Fred. No. 0.67;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

144 AGCCACCAAGTCGACAGCCCTCACGGCAGGGCTGGGTGTACACAGCTACCTGGACAGGA 203
      |||||
412 AGCCCGAGGACGACGCGGTTCAGGACAGGGCTGGATGGCCCGCCCCCACCAGTGGAGG 471
      |||||

294 GGCTCAAGCTGTCACTGAGTGGGGGCCGCTGAGGCCCTCTGAGTTCCCTGGGGAGGCTG 263
      |||||
472 AGCTGAAGCGCGCTGGAGCGCGGGCAGCTGGAGGCGCGGGCGCTGTCTGGCGCTGG 531
      |||||

264 TGTCTGAGACGAATACAGGCAGGCTGAGCTCTGTGATCCAGGAGCTGCTGAGTTCTG 323
      |||||
532 AGCGGGAGCTGGCGGGCGGCGGCGCGCGGCTGTGACGAGGAGGAGCTGTGTGCGGCG 591
      |||||

324 AGCAGGCTTTCGTGGAGGAGCTGCAGTTCTCTGCAGAGCCACACCTGCAGCACCTGGAGC 383
      |||||
592 GCCAGAGCAAGTGTGAGGCGCTGTACGAGTGTGTGCGCACCAAGTGTGTGGCGCTGCTGC 651
      |||||

384 G 384
      |
652 G 652

RESULT 9
AAC84312
ID AAC84312 standard; cDNA; 1803 BP.
XX

```

Query Match 10.8%; Score 41.4; DB 5; Length 1803;
 Best Local Similarity 47.5%; Pred. No. 1.1;
 Matches 123; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 115 CCACCCACTGCGCTGCTGTCGACCAAGCCCAAGTCCAGCCCTCAGCGGAGGG 174
 DB 1501 CCAAGACTGGCTGGGTGGCGAGGGAGCCAGATCCCGAGGGAGGACCTCGAGGG 1560

QY 175 CTGGGTGTCAACAGCCTACCTGGACAGGAGCTCAAGCTGTCACTGAGTGGGGGGCGC 234
 DB 1561 CCGCAAGATCGAGCCCCCAGCTGGGAGGGGAGGCCGGTGCCTCAGGGGGCGTGGC 1620

QY 235 TGAGCCCTCAGTTCCCTGGGAGGCTGTCTTGAGAGCAATACAGGCAAGGCTGAG 294
 DB 1621 ACACTGCCCTTCCCGACGGGTGGCAGGCCCTGGAGGAACTGAGTGTCACTGTAT 1680

QY 295 CTCGTGTATCAGGAGCTGCTGAGTCTTGAGCAGCCCTTGTGGAGGAGCTGAGTTCT 354
 DB 1681 CTCAGGCACAGCTCTGCGGGCTCCAGCCGGCTCTGAGAGCCGCTGAAAGTCA 1740

QY 355 CGAGAGCCACCACTGCAG 373
 DB 1741 GCGACTTAAGCTTGACAG 1759

RESULT 10

ACA23561

ID ACA23561 standard; DNA; 1278 BP.

XX AC

ACA23561;

XX AC

DT 19-JUN-2003 (first entry)

XX DE

Prokaryotic essential gene #5218.

XX KW

Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KW

drug design; gene.

XX OS

Borrelia cepacia.

XX XX

PN W020027183-A2.

XX PD

03-OCT-2002.

XX PF

21-MAR-2002; 2002WO-US009107.

XX PR

21-MAR-2001; 2001US-00815242.

XX PR

06-SEP-2001; 2001US-00948993.

XX PR

25-OCT-2001; 2001US-0342923P.

XX PR

08-FEB-2002; 2002US-00072851.

XX PR

06-MAR-2002; 2002US-0362699P.

XX XX

(ELIT-) ELITRA PHARM INC.

XX PA

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW;

XX PI

Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX XX

WPI; 2003-029926/02.

XX DR

P-PSDB; ABU19691.

XX PT

New antisense nucleic acids, useful for identifying proteins or screening

XX PT

for homologous nucleic acids required for cellular proliferation to

XX PT

isolate candidate molecules for rational drug discovery programs.

XX PS

Claim 14; SEQ ID NO 11431; 1766pp; English.

XX CC

The invention relates to an isolated nucleic acid comprising any one of

CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 1278 BP; 199 A; 454 C; 440 G; 185 T; 0 U; 0 Other;

Query Match

Best Local Similarity 10.3%; Score 39.8; DB 7; Length 1278;

Matches 161; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 5 GCGGAGATCTTGACATCTAGCTGCTCAGCTGCTACCTGCTCCCTAGGGCTGACGAG 64

DB 655 GCGGTGCGCTATCAGCTGAGCTGCCGAAACGCGCCACAAAGCGGGCTTCTAGCTGACCCAG 714

QY 65 GATGCATCAGCTCGGGGAGGCCAGTATGTGGAGGTCTCTGGATGCAGCCACCCACTG 124

DB 715 CGCGACAACCGCGCGCTCTGCGCGAGTACGCGGCCGACCGAGCTGCTGAACCTGCTTC 774

QY 125 CGCTGGCTTTCGCGACCAAGCCCAAGTCCAGCCCTCAGCGCAGGGTGGGTGTC 184

DB 775 TGCTACACGGCGGCTTCTGCTCGCGCGCTCAAGGCGCGCGGAAACGCGTCTGTGTCG 834

QY 185 CCAGCTACCTGGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGGCGCTGAGGCCCT 244

DB 835 ATCGACTCGTGGGCGATGCGCTCGCGTGGGAGGAGACGCTGTCGCGCAACGGCTTC 894

QY 245 GAGTTCCTCGGGAGGCTGTGTGTAAGACGAAATACAAAGGCAAGGCTGAGTCTGTGATC 304

DB 895 GACGCCGAACCGCGACACCTGGCTCGACGCCGATGCGTTCAAGACGCTGCGCGCTGTC 954

QY 305 CAGGAGCTGTGAGTTCGAGCAGGCCCTTCGTGGAGGAGCTGCAGTTCCTGCAGAGCCAC 364

DB 955 GACGAGGAGCAACGATTCGACTGATGCTGATCGCTCGCGGAGGCTGCGCCGACCGC 1014

QY 365 CAC 367

DB 1015 GAC 1017

RESULT 11

ACH34588

ID ACH34588 standard; cDNA; 455 BP.

XX AC

ACH34588;

XX AC

DT 13-OCT-2003 (first entry)

XX DE

Human endothelial cell cDNA #2721.

XX KW

Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST; genome mapping; biodiversity; genetic disorder.

OS Homo sapiens.
XX US2003073623-A1.
XX 17-APR-2003.
XX 30-JUL-2001; 2001US-00918995.
XX 30-JUL-2001; 2001US-00918995.
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX (STAC/) STACHE-CRAIN B.
XX (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX New polynucleotide sequences obtained from various cDNA libraries, useful
XX as hybridization probes, as oligomers for PCR, for chromosome and gene
XX mapping, in the recombinant production of protein, or in generating
XX antisense DNA or RNA.
XX Claim 1; SEQ ID NO 21800; 44pp; English.
XX The invention relates to an isolated polynucleotide comprising any one of
XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
XX determined by the technique of SBH (sequencing by hybridisation). Also
XX included is a purified polypeptide comprising a sequence corresponding to
XX a reading frame of the novel polynucleotide. The nucleic acid sequences
XX are useful in diagnostics as expressed sequence tags (EST) for
XX identifying expressed genes or for physical mapping of the human genome,
XX in forensics, in assessing biodiversity, or in identifying mutations
XX responsible for genetic disorders and other traits. The nucleotide
XX sequences are also useful as hybridisation probes, as oligomers for PCR,
XX for chromosome and gene mapping, in the recombinant production of
XX protein, or in generating antisense DNA or RNA. The purified polypeptide
XX is useful for generating antibodies specific for it. The present sequence
XX is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=20030073623
XX SQ Sequence 455 BP; 81 A; 140 C; 143 G; 90 T; 0 U; 1 Other;
Query Match 10.2%; Score 39.4; DB 8; Length 455;
Best Local Similarity 51.4%; Pred. No. 2.6;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 18 ACATCTACGTGTCACCGCTGACTACTCGCCCTAGGGGCTGAGCAGGATGCCATCAGC 77
DB 91 AGCCCTTCGTGAGCAGCGCTAAGTCCAGCTCTCTTTGCCCTCAAGTGTCTCAACATGA 150
QY 78 TCGGGGAAGGCCAGTATGTGAGGTCCTGATGAGCCACCCACCTGCGCTGGCTTGTCC 137
DB 151 TGCCCGAGGAGAAAGCTGTGAGGCGCTTGGTGCAGCCACCGAGAGAGAGAGGCC 210
QY 138 GCACCAAGCCCAAGTCCAGCCCTCAGCGAGGGCTGGGTGTCCAGCAGCTTACC 194
DB 211 TGGAGAAGTTGCTCCCGCCTCTCTTGAGGGCTGCTGAGCTGGTGGCACCCTCCC 267
RESULT 12
ID ACH34675
XX ACH34675 standard; cDNA; 474 BP.
XX ACH34675;
XX 13-OCT-2003 (first entry)
XX Human endothelial cell cDNA #2808.

XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX Homo sapiens.
XX US2003073623-A1.
XX 17-APR-2003.
XX 30-JUL-2001; 2001US-00918995.
XX 30-JUL-2001; 2001US-00918995.
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX (STAC/) STACHE-CRAIN B.
XX (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX New polynucleotide sequences obtained from various cDNA libraries, useful
XX as hybridization probes, as oligomers for PCR, for chromosome and gene
XX mapping, in the recombinant production of protein, or in generating
XX antisense DNA or RNA.
XX Claim 1; SEQ ID NO 21887; 44pp; English.
XX The invention relates to an isolated polynucleotide comprising any one of
XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
XX determined by the technique of SBH (sequencing by hybridisation). Also
XX included is a purified polypeptide comprising a sequence corresponding to
XX a reading frame of the novel polynucleotide. The nucleic acid sequences
XX are useful in diagnostics as expressed sequence tags (EST) for
XX identifying expressed genes or for physical mapping of the human genome,
XX in forensics, in assessing biodiversity, or in identifying mutations
XX responsible for genetic disorders and other traits. The nucleotide
XX sequences are also useful as hybridisation probes, as oligomers for PCR,
XX for chromosome and gene mapping, in the recombinant production of
XX protein, or in generating antisense DNA or RNA. The purified polypeptide
XX is useful for generating antibodies specific for it. The present sequence
XX is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=20030073623
XX SQ Sequence 474 BP; 83 A; 150 C; 147 G; 94 T; 0 U; 0 Other;
Query Match 10.2%; Score 39.4; DB 8; Length 474;
Best Local Similarity 51.4%; Pred. No. 2.6;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 18 ACATCTACGTGTCACCGCTGACTACTCGCCCTAGGGGCTGAGCAGGATGCCATCAGC 77
DB 92 AGCCCTTCGTGAGCAGCGCTAAGTCCAGCTCTCTTTGCCCTCAAGTGTCTCAACATGA 151
QY 78 TCGGGGAAGCCAGTATGTGAGGTCCTGATGAGCCACCCACCTGCGCTGGCTTGTCC 137
DB 152 TGCCCGAGGAGAAAGCTGTGAGGCGCTTGGTGCAGCCACCGAGAGAGAGAGGCC 211
QY 138 GCACCAAGCCCAAGTCCAGCCCTCAGCGAGGGCTGGGTGTCCAGCAGCTTACC 194
DB 212 TGGAGAAGTTGCTCCCGCCTCTCTTGAGGGCTGCTGAGCTGGTGGCACCCTCCC 268
RESULT 13
ID ABS55941
XX ABS55941 standard; cDNA; 1302 BP.
XX AC ABS55941;

XX 22-JAN-2003 (first entry)
XX DNA topoisomerase II (TOP2) 21.34 cDNA.
XX DNA topoisomerase II 21.34; TOP2; gene; ss; DNA recombination; cancer;
KW malignant tumour; haemopathy; human immunodeficiency virus; HIV;
KW immunological disease; inflammation; development disturbance.
XX Unidentified.
OS
XX
XX Key Location/Qualifiers
PH 48. .632
FT CDS /tag= a
FT /product= "DNA topoisomerase II 21.34"
XX
XX CN1345941-A.
XX
XX 24-APR-2002.
XX
XX 29-SEP-2000; 2000CN-00125577.
XX
XX 29-SEP-2000; 2000CN-00125577.
XX
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX
XX Mao Y, Xie Y;
XX
XX WPI: 2002-539340/58.
XX P-PSDB; ASG70542.
XX
XX New polypeptide-DNA topoisomerase II (TOP2) 21.34 for treating malignant
PT tumor, hemopathy, development disturbance, human immunodeficiency virus
PT infection, immunological disease and various inflammations.
XX
XX Claim 6; Page 26-27 (Disclosure); 34pp; Chinese.
XX
XX The invention relates to the polypeptide DNA topoisomerase II (TOP2)
CC 21.34, a polynucleotide encoding the polypeptide and a method for
CC producing the polypeptide by DNA recombination technology. The
CC polypeptide is used for curing several diseases, such as malignant
CC tumours, haemopathy, development disturbance, human immunodeficiency
CC virus (HIV) infection, immunological diseases and various inflammations.
CC This sequence represents cDNA encoding DNA topoisomerase II (TOP2) 21.34
XX
SQ Sequence 1302 BP; 216 A; 391 C; 422 G; 273 T; 0 U; 0 Other;
Query Match 10.2%; Score 39.4; DB 6; Length 1302;
Best Local Similarity 51.4%; Pred. NO. 3.3;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 18 ACATCTAGTGTCTACCCCTGACTCTGCCCCCTAGGGGCTGACGAGTCCATCAGC 77
Db 484 ACGCTTCTGTCGAGCAGGCTAAGTGCCAGCTCTTTTGGCCCTCAAGGTGCTCAACATGA 543
QY 78 TCGCGGAAGGCGCAGTATGTGAGGTCTCTGATGCGAGCCACCCACTGCGTGTGCTTCC 137
Db 544 TGCCCGAGGAGAGTGTGTTGAGCCCTGCTGCGCCACCGAGAGCAGAGAGGCC 603
QY 138 GCACCAAGCCACCAAGTTCAGCCCTCAGCGAGGGCTGGGTGTACGAGCTTACC 194
Db 604 TGGAGAAGTGTCTCCCGGCTCTTGTAGGGGCTGCTGAGCTGGTGGCCCTCC 660
RESULT 14
AAS26907
ID AAS26907 standard; cDNA; 1354 BP.
XX
XX AAS26907;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human cDNA encoding a novel secreted protein, SEQ ID 99.

XX KW Human; immunosuppressive; antiarthritic; ss; antitumour; cytostatic;
XX KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
XX KW antibacterial; virucide; fungicide; ophthalmological; vulnary;
XX KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
XX KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
XX KW cerebral ischaemia; angiogenesis; nervous system disorder;
XX KW Alzheimer's disease; infection; ocular disorder; corneal infection;
XX KW wound healing; epithelial cell proliferation; skin ageing; food additive;
XX KW preservative; antiproliferative.
XX
XX Homo sapiens.
XX
XX WO200155441-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001320.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226686P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229503P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232080P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.

PR 14-SEP-2000; 2000US-02332399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239933P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0245207P.
PR 17-NOV-2000; 2000US-0245208P.
PR 17-NOV-2000; 2000US-0245209P.
PR 17-NOV-2000; 2000US-0245210P.
PR 17-NOV-2000; 2000US-0245211P.
PR 17-NOV-2000; 2000US-0245212P.
PR 17-NOV-2000; 2000US-0245213P.
PR 17-NOV-2000; 2000US-0245214P.
PR 17-NOV-2000; 2000US-0245215P.
PR 17-NOV-2000; 2000US-0245216P.
PR 17-NOV-2000; 2000US-0245217P.
PR 17-NOV-2000; 2000US-0245218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0245264P.
PR 17-NOV-2000; 2000US-0245265P.
PR 17-NOV-2000; 2000US-0245297P.
PR 17-NOV-2000; 2000US-0245299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.

PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476222/51.
DR P-PSDB; AAU17002.
XX
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, for treating blood clotting disorder,
PT hemophilia.
XX
XX Claim 1; SEQ ID NO 99; 601pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention. Note: The
CC
Query Match 10.2%; Score 39.4; DB 4; Length 1354;
Best Local Similarity 51.4%; Pred. No. 3.3;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 18 ACATCTACGTGTCACCGCTGACTACCTGCCCTAGGGGCTGAGGAGTGCATCAGC 77
Db 511 ACGCCTTCGTGAGCAGCGCTAAGTGCAGCTCCTCTTTGCCCTCAAGTGCATGA 570
QY 78 TGGGGAAGCCAGTATGAGGAGTCTGGATGCGAGCCACCCACTGCGCTGCTTCC 137
Db 571 TGCCGAGGAGAGCTGGTTGAGGCCCTGCTGCGAGCCACCAGAGGAGAGAGGCC 630
QY 138 GCACCAAGCCACCAAGTCCAGCCCTCAGCGAGGGCTGGGTGTACACAGCTTACC 194
Db 631 TGGAGAAGTTGCTCCCGGCTCTCTTTGAGGGCTGCTGAGCTGGTGGCACCTGCC 687
RESULT 15
ABQ54643
ID ABQ54643 standard; cDNA; 1354 BP.
XX
AC ABQ54643;
XX
DT 22-AUG-2002 (first entry)
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:09:24 ; Search time 87.2015 Seconds
(without alignments)
6179.453 Million cell updates/sec

Title: US-10-077-130-4_COPY_23150_24120

Perfect score: 971

Sequence: 1 cagcggcggtgctgctggcg.....ggcgctaaaaaagtctaga 971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 926.8 | 95.4 | 5207 | 4 | US-09-858-664A-1 |
| 2 | 926.8 | 95.4 | 5207 | 4 | US-10-274-978-1 |
| 3 | 926.8 | 95.4 | 5207 | 4 | US-10-274-978-3 |
| 4 | 112.2 | 11.6 | 2132 | 2 | US-09-159-385-3 |
| 5 | 112.2 | 11.6 | 2132 | 2 | US-09-186-277-3 |
| 6 | 94.2 | 9.7 | 1429 | 2 | US-09-159-385-4 |
| 7 | 94.2 | 9.7 | 1429 | 3 | US-09-186-277-4 |
| 8 | 83.2 | 8.6 | 1282 | 2 | US-08-878-989-12 |
| 9 | 83.2 | 8.6 | 1282 | 3 | US-09-272-796-12 |
| 10 | 83.2 | 8.6 | 1282 | 4 | US-09-016-434-953 |
| 11 | 76.6 | 7.9 | 1458 | 4 | US-09-230-896C-5 |
| 12 | 71.2 | 7.3 | 2169 | 4 | US-09-016-434-1147 |
| 13 | 69.6 | 7.2 | 4935 | 2 | US-08-631-087-3 |
| 14 | 69.6 | 7.2 | 5886 | 3 | US-08-810-712-9 |
| 15 | 69.6 | 7.1 | 3228 | 4 | US-09-579-664B-5 |
| 16 | 68.6 | 7.1 | 678 | 4 | US-09-841-683-4 |
| 17 | 68.6 | 7.1 | 711 | 4 | US-09-841-683-6 |
| 18 | 68.6 | 7.1 | 1191 | 4 | US-09-841-683-10 |
| 19 | 68.6 | 7.1 | 1224 | 4 | US-09-841-683-8 |
| 20 | 68.6 | 7.1 | 1485 | 4 | US-09-801-876B-1 |
| 21 | 68.6 | 7.1 | 1485 | 4 | US-10-254-869-1 |
| 22 | 68.6 | 7.1 | 1675 | 4 | US-09-841-683-12 |
| 23 | 68.2 | 7.0 | 1349 | 1 | US-07-951-715A-20 |
| 24 | 68.2 | 7.0 | 1349 | 2 | US-08-459-448A-20 |
| 25 | 68.2 | 7.0 | 1349 | 3 | US-08-459-595A-20 |
| 26 | 68.2 | 7.0 | 1349 | 3 | US-08-459-504B-20 |
| 27 | 68.2 | 7.0 | 1349 | 3 | US-08-459-444-20 |

ALIGNMENTS

RESULT 1

US-09-858-664A-1
; Sequence 1, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-858-664A-1

| | | | | |
|-----------------------|-------|---|--------|---------------|
| Query Match | 95.4% | Score 926.8; | DB 4; | Length 5207; |
| Best Local Similarity | 99.8% | Pred. No. 1.1e-184; | | |
| Matches | 928; | Conservative | 0; | Mismatches 2; |
| | | | Indels | 0; |
| | | | Gaps | 0; |
| Qy | 1 | CAGCGGGGGGGCTGGCGGCCCAAGATCATCCCTACCAACCAAGACACAGCAGT | 60 | |
| Db | 4278 | CAGCGGGGGGGCTGGCGGCCCAAGATCATCCCTACCAACCAAGACACAGCAGT | 4337 | |
| Qy | 61 | GCTCGCGGAATACAGGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCCGACGCTG | 120 | |
| Db | 4338 | GCTCGCGGAATACAGGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCCGACGCTG | 4397 | |
| Qy | 121 | AGCTTACCTCAGCCCCCGGCACTGGTGTCTATCTGGAGCTGTGTCTGGGCCCCAGCT | 180 | |
| Db | 4398 | AGCTTACCTCAGCCCCCGGCACTGGTGTCTATCTGGAGCTGTGTCTGGGCCCCAGCT | 4457 | |
| Qy | 181 | GCTCCCTGCTGGCGAGAGGGCTTCTACTCAGAAATCGAGGTGAAGACTACCTGTG | 240 | |
| Db | 4458 | GCTCCCTGCTGGCGAGAGGGCTTCTACTCAGAAATCGAGGTGAAGACTACCTGTG | 4517 | |
| Qy | 241 | GCAGATGTTGAGTGCCACCCAGTACCTGCAACCAAGACATCTCTCAGCTGACCTGAG | 300 | |
| Db | 4518 | GCAGATGTTGAGTGCCACCCAGTACCTGCAACCAAGACATCTCTCAGCTGACCTGAG | 4577 | |
| Qy | 301 | GTCGAGAACATGATCATACCCAAATACACCTGCTCAGGTGCTGGACCTGGGCAATGC | 360 | |
| Db | 4578 | GTCGAGAACATGATCATACCCAAATACACCTGCTCAGGTGCTGGACCTGGGCAATGC | 4637 | |

| | | |
|-------|---------------------------|---|
| / | PRIOR APPLICATION NUMBER: | 09/858,664 |
| / | PRIOR FILING DATE: | 2001-05-17 |
| / | PRIOR APPLICATION NUMBER: | 09/711,134 |
| / | PRIOR FILING DATE: | 2000-11-14 |
| / | NUMBER OF SEQ ID NOS: | 34 |
| / | SOFTWARE: | FastSeq for Windows Version 4.0 |
| / | SEQ ID NO 3 | |
| / | LENGTH: | 5207 |
| / | TYPE: | DNA |
| / | ORGANISM: | Human |
| / | US-10-774-978--3 | |
| <hr/> | | |
| | Query Match | 95.4%; Score 926.8; DB 4; Length 5207; |
| | Best Local Similarity | 99.8%; Pred. No. 1.le-184; |
| | Matches 928; Conservative | 0; Mismatches 2; Indels 0; Gaps 0; |
| Qy | 1 | CAGCGGGCGGCGCTGGCCGCCAAGAATCATCCCTTACCACCCCACAAGGACAAGACACGACT 60 |
| Db | 4278 | CAGCGGGCGGCGCTGGCCGCCAAGAATCATCCCTTACCACCCCACAAGGACAAGACACGACT 4337 |
| Qy | 61 | GCTGCGCGGAATACGAGGCCTCAAGGCCTGCGCCACCCGACCTGGCCAGCTGCACGC 120 |
| Db | 4338 | GCTGCGCGGAATACGAGGCCTCAAGGCCTGCGCCACCCGACCTGGCCAGCTGCACGC 4397 |
| Qy | 121 | AGCCTACTCAGCCCCCGCACCTGGTGTCATCTTTGAGCTGTGTCTGGGCCCGAGCT 180 |
| Db | 4398 | AGCCTACTCAGCCCCCGCACCTGGTGTCATCTTTGAGCTGTGTCTGGGCCCGAGCT 4457 |
| Qy | 181 | GCTCCCTGCTGGCGAGAGGGCCTCTACTCAGAACTCGAGGTGAAGACTACCTGTG 240 |
| Db | 4458 | GCTCCCTGCTGGCGAGAGGGCCTCTACTCAGAACTCGAGGTGAAGACTACCTGTG 4517 |
| Qy | 241 | GCAGATTTCAGTAGGCCACCCACGACTCTGCAACACCCAGCACATCTGCACCTGGACCTGAG 300 |
| Db | 4518 | GCAGATTTCAGTAGGCCACCCACGACTCTGCAACACCCAGCACATCTGCACCTGGACCTGAG 4577 |
| Qy | 301 | GTCCGAGAACATGATCATACCGAATAACAACCTGTCTCAAGGTGTTGACCTGGCAATGC 360 |
| Db | 4578 | GTCCGAGAACATGATCATACCGAATAACAACCTGTCTCAAGGTGTTGACCTGGCAATGC 4637 |
| Qy | 361 | ACAGAGCTCAGCCAGAGAGGTGTCGCCCTCAGACAAGTTCAAGGACTACCTAGAGAC 420 |
| Db | 4638 | ACAGAGCTCAGCCAGAGAGGTGTCGCCCTCAGACAAGTTCAAGGACTACCTAGAGAC 4697 |
| Qy | 421 | CATGGCTCCAGAGCTCTCTGAGGGCCAGGGGGCTGTTCCAAGACAGACATCTGGGCCAT 480 |
| Db | 4698 | CATGGCTCCAGAGCTCTCTGAGGGCCAGGGGGCTGTTCCAAGACAGACATCTGGGCCAT 4757 |
| Qy | 481 | CGGTGTACAGCCTTTCATCATGTGAGCGCGAGTACCCGCTGAGCAGCAGAGGTGCACG 540 |
| Db | 4758 | CGGTGTACAGCCTTTCATCATGTGAGCGCGAGTACCCGCTGAGCAGCAGAGGTGCACG 4817 |
| Qy | 541 | CGACCTTCAGAGAGSACTGCGCAAGGGCTGTCTCGGCTGAGCGGTGTCTACCGGGGCT 600 |
| Db | 4818 | CGACCTTCAGAGAGSACTGCGCAAGGGCTGTCTCGGCTGAGCGGTGTCTACCGGGGCT 4877 |
| Qy | 601 | GTCCGGGGCGCGCTGGCCCTCTCTGCGACATCTGTGCGCCAGCCCTGGGGCGGGCC 660 |
| Db | 4878 | GTCCGGGGCGCGCTGGCCCTCTCTGCGACATCTGTGCGCCAGCCCTGGGGCGGGCC 4937 |
| Qy | 661 | CTGCGCGTCCAGCTGCTTCAGTGCCTGCTGAGTGCCTGCTTAACAGAGAGGGCCCGGCTGTTCCGCG 720 |
| Db | 4938 | CTGCGCGTCCAGCTGCTTCAGTGCCTGAGTGCCTGCTTAACAGAGAGGGCCCGGCTGTTCCGCG 4997 |
| Qy | 721 | GCCCGCGCCGTGACCTTTCCTTACCGCGCGGCTCGCGGTCTTTCGTGCGCAATCGCAGAA 780 |
| Db | 4998 | GCCCGCGCCGTGACCTTTCCTTACCGCGCGGCTCGCGGTCTTTCGTGCGCAATCGCAGAA 5057 |
| Qy | 781 | GAGACGGCGCTGCTGTATACAGAGGCACAACCTGGCCCGAGGTGGCTGAGGGTGC GCCCG 840 |
| Db | 5058 | GAGACGGCGCTGCTGTATACAGAGGCACAACCTGGCCCGAGGTGGCTGAGGGTGC GCCCG 5117 |
| Qy | 841 | GCACACCCCTTGGTCTCCCGCTGGGGGTGCTCTGACAGCGCGCCAAATAAAAACGCC CAGC 900 |

Db 5118 GCACACCCCTTGCTCTCCCGCTGGGGTGCCTGCAGACGGCCCAATAAAAAACGCACAGC 5177

Qy 901 CGCGCGAGAAAAAATAAAAAA 930

Db 5178 CGGCGCAGAAAAAATAAAAAA 5207

RESULT 4

US-09-159-385-3

; Sequence 3, Application US/09159385

; Patent No. 5958748

; GENERAL INFORMATION:

; APPLICANT: AKIRA, SHIZUO

; APPLICANT: KAWAI, TARO

; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE

; FILE REFERENCE: PH-569

; CURRENT APPLICATION NUMBER: US/09/159,385

; CURRENT FILING DATE: 1998-09-23

; EARLIER APPLICATION NUMBER: JP97/261589

; EARLIER FILING DATE: 1997-09-26

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 2132

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (94)..(1455)

US-09-159-385-3

Query Match 11.6%; Score 112.2; DB 2; Length 2132;

Best Local Similarity 53.9%; Pred. No. 1.2e-14;

Matches 256; Conservative 0; Mismatches 213; Indels 6; Gaps 1;

Qy 64 GCGCGATACGAGCCCTCAAGGCGCTGCGCCACCGCACCTGCCAGCTGCCAGCGCAGC 123

Db 279 GCGGAGGTGAACAATCTCGCGGAGATCCGGCACCCCAACATCATCACCCTGCAGCAT 338

Qy 124 CTACCTCAGCCCCCGGCACCTGGTGTCTCATCTTGGAGCTGTGCTCTGGGCCGAGGTGCT 183

Db 339 CTTGAGAACAAAGACGACGTGGTCTCATCTTGAGCTGTGCTCTGGCGGAGCTCTT 398

Qy 184 CCGCTGCTGCGCGAGAGGGCTCTACTCAGANTCGAGGTGAAGGACTACCTGTGGCA 243

Db 399 TGACTTCTCGCGGAGAAAGAGTGCCTGACGAGGACGAGGCCACCCAGTTCCTCAAGCA 458

Qy 244 GATGTTGAGTGCACCCACAGTACCTGTCACACACGACATCTGACCTGGACCTGAGTGC 303

Db 459 GATCTTGAGCGGGTTCACATCTGCTGCTCTAAGCGCATCGCACATTTGACCTGAAGCC 518

Qy 304 CGAAGCATGATCATCCGAATACACCTGCTCAA-----GTCGTGACCTGGGCAA 357

Db 519 GGAAAAATCATGCTGCTGAGCAAGAACGTGCCCAACCCACGAATCAAGCTCATCGACTT 578

Qy 358 TGCACAGAGCCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAGTTCAAGGACTACCTAGA 417

Db 579 CGGCATCGCGCAAGAATCGAGCGGGAGACGAGTTCAAGAACATCTTCGCACCCCGGA 638

Qy 418 GACCATGGCTCCAGAGTCTCTGAGAGGCGCAGGGGGCTGTTCCACAGACAGCATCTGGGC 477

Db 639 GTTTGTGGCCCCAGAGATTGTGAACATATGAGCCGCTGGGGCTGGAGCGGACATGTGGAG 698

Qy 478 CATCGGTGTGACAGCCCTCATCATGCTGAGCGCGGAGTACCCGCTGAGCAGCGAG 532

Db 699 CATCGGTGTATCATCACTATATCTCTCTGAGCGGTGCATCCCGCTTCTGGGCGAG 753

RESULT 5

US-09-186-277-3

; Sequence 3, Application US/09186277

; Patent No. 6171841

RESULT 5
US-09-186-277-3
; Sequence 3, Application US/09186277
; Patent No. 6171841

; GENERAL INFORMATION:
 ; APPLICANT: AKIRA, SHIZUO
 ; APPLICANT: KAWAI, TARO
 ; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE
 ; FILE REFERENCE: 081356/0128
 ; CURRENT APPLICATION NUMBER: US/09/186,277
 ; CURRENT FILING DATE: 1998-11-05
 ; EARLIER APPLICATION NUMBER: JP97/261589
 ; EARLIER FILING DATE: 1997-09-26
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 2132
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (94)..(1455)
 US-09-186-277-3

Query Match 11.6%; Score 112.2; DB 3; Length 2132;
 Best Local Similarity 53.9%; Pred. No. 1.2e-14;
 Matches 256; Conservative 0; Mismatches 213; Indels 6; Gaps 1;
 QY 64 GCGGAATACAGAGGCTCAAGGGCTCGGCCACCGCAGCTGGCCAGCTGACGAGC 123
 DB 279 GCGGAGGTGAACATCTCTCGGGAGATCCGGACCCCAACATCATCACCCTGACGACAT 338
 QY 124 CTACCTAGCCCCCGACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCGGAGCTGCT 183
 DB 339 CTTGAGAAACAAGACGCGTGGTCTCATCTCTGGAGCTGTCTCTGGCGGGAGCTCTT 398
 QY 184 CCCCTGCTGCGCAGAGGGCTCTACTACAGATCCGAGTGAAGGACTACCTGTGGCA 243
 DB 399 TGACTTCTTGGGAGAAAGAGTGGCTGACGAGAGGAGGCGACCCAGTTCTCTCAAGCA 458
 QY 244 GATGTTGAGTGCCACCCAGTACCTGCAACACGACATCTCTGACCTGGAAGCTGAGGTC 303
 DB 459 GATCTGACGCGGTCTACTACCTGCACTTAAGCGCATCGACACTTGAACCTGAAGCC 518
 QY 304 CGAGACATGATCATACCGATACACCTGCTCAA-----GGTGTGACCTGGGCAA 357
 DB 519 GGAATAACATCATGCTGTGGCAAGAACGTCGCCCAACCCAGATCAAGCTCATCGACTT 578
 QY 358 TGCACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGA 417
 DB 579 CGGCATCGCACAGATCGAGGCGGGGAACGAGTTCAAGAACATCTTCGGCACCCCGGA 638
 QY 418 GACATGCTCAGAGCTCTGGAGGGCGAGGGGCTGTTCCACAGACAGACATCTTGGGC 477
 DB 639 GTTGTGGCCCGCAGAGATTGTAATATGAGCCCTGGGCTGAGGCGGACATGTGGAG 698
 QY 478 CATGGTGTACAGCTTCATCATGCTGAGCGCGAGTACCCTGGTGAAGCGGAG 532
 DB 599 CATGGTGTACATCATATATCTCTGAGCGGTGCAATCCCGGTTCTCTGGGCGAG 753

RESULT 6
 US-09-159-385-4
 ; Sequence 4, Application US/09159385
 ; Patent No. 5958748
 ; GENERAL INFORMATION:
 ; APPLICANT: AKIRA, SHIZUO
 ; APPLICANT: KAWAI, TARO
 ; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE
 ; FILE REFERENCE: PH-569
 ; CURRENT APPLICATION NUMBER: US/09/159,385
 ; CURRENT FILING DATE: 1998-09-23
 ; EARLIER APPLICATION NUMBER: JP97/261589
 ; EARLIER FILING DATE: 1997-09-26
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4

; LENGTH: 1429
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (10)..(1353)
 US-09-159-385-4
 Query Match 9.7%; Score 94.2; DB 2; Length 1429;
 Best Local Similarity 52.3%; Pred. No. 6.2e-11;
 Matches 235; Conservative 0; Mismatches 208; Indels 6; Gaps 1;
 QY 90 TGGGCCACCGCAGCTGGCCAGCTGACGAGCCTACTCAGCCCGGCGACCTGGTGC 149
 DB 221 TCGGCCACCCCAACATCATACATGTCATGCTTCGAGAACAGACAGATGTGGTGC 280
 QY 150 TCATCTTGGAGCTGTGCTTGGGCCCGAGCTGCTCCCTGCTGGCCGAGAGGGCTCCT 209
 DB 281 TGTCTCTGGAGCTGGTCCGGTGGGAGCTTTTCGACTTCTGGCCGAGAGGAGTCA 340
 QY 210 ACTCAGATCCGAGTGAAGACTACCTGTGGCAGATGTTGAGTGCCACCCAGTACCTGC 269
 DB 341 TGACGGAGGATGAGGCCACGACAGTTCTCAAAACAAATCTAGACGGTGTCCACTACCTGC 400
 QY 270 ACAACCGACACATCTGACCTGGACCTGAGTCCGAGAACATGATCATCACCAGATACA 329
 DB 401 ACTCAAGCGCATCGACACTTTGACCTGAAGCCGAGAACATCATGTTGCTGGCAAGC 460
 QY 330 A-----CTGTCTAAGTGTGGACCTGGGCAATGCAAGAGCTCAGCCAGGAGAGG 383
 DB 461 ACCGACGAGCGCCCGCATTAAGCTCATCGACTTTGGCATCGCGCACAGGATCGAGGCTG 520
 QY 384 TGCTGCCCTCAGACAGTTCAGAGACTACCTAGAGACCATGCTCCAGAGCTCCTGGAGG 443
 DB 521 GGAGCGAGTTCAGAACATCTTTGGCACACCCGAGTTTGTGCCCCCGAGATCGTGAAT 580
 QY 444 GCCAGGGGCTGTTCCACAGACAGACATCTGGGCCATCGGTGTGACAGCTTCATCATGC 503
 DB 581 ATGAGCCACTTGGCTTGGAGGCTGACATGTGAGCATTTGGCGTATCATCATCATCTCC 640
 QY 504 TGAGCGCGAGTACCCGCTGAGCAGCGAG 532
 DB 641 TGAGCGGAGCGTCCCATTCCTGGGCGAG 669

RESULT 7
 US-09-186-277-4
 ; Sequence 4, Application US/09186277
 ; Patent No. 6171841
 ; GENERAL INFORMATION:
 ; APPLICANT: AKIRA, SHIZUO
 ; APPLICANT: KAWAI, TARO
 ; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE
 ; FILE REFERENCE: 081356/0128
 ; CURRENT APPLICATION NUMBER: US/09/186,277
 ; CURRENT FILING DATE: 1998-11-05
 ; EARLIER APPLICATION NUMBER: JP97/261589
 ; EARLIER FILING DATE: 1997-09-26
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 1429
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (10)..(1353)
 US-09-186-277-4

Query Match 9.7%; Score 94.2; DB 3; Length 1429;
 Best Local Similarity 52.3%; Pred. No. 6.2e-11;
 Matches 235; Conservative 0; Mismatches 208; Indels 6; Gaps 1;
 ; SEQ ID NO 4

QY 90 TGGCCACCCGACCTGGCCGAGCTGACGAGCTACCTCAGCCCGGACCTGGTGC 149
Db 221 TCGCCACCCCAACATCATACACTGCATGAGCTTTCGAAACAAGACAGATGGTGC 280
QY 150 TCATCTTGGAGCTGTCTCTGGGCCGAGCTGCTCCCTGCTGGCCGAGAGGGCTCT 209
Db 281 TGATCCTGGAGCTGTCTCGGTGGCGAGCTTTTCGACTTCTTGGCCGAGAAGAGTCA 340
QY 210 ACTCAGATCCGAGTGAAGACTACTCTGTGGCAGATGTTGAGTGCCACCCAGTACCTGC 269
Db 341 TGACGAGGATAGGCCACGACAGTTCCTCAACAATCCTAGACGGTGTCCACACTCTGC 400
QY 270 ACAACGACACATCTCTGACCTGGACCTGAGCTCCGAGAATGATCATCAACCAATACA 329
Db 401 ACTCAAGGCGATCGACACATTTGACCTGAAGCCGAGACATCATGTTGTGGACAAGC 460
QY 330 A-----CCTGCTAAGGTGCTGGACCTGGGCAATGACAGACCTCAGCCAGAGAGG 383
Db 461 ACGACGACGAGCCCGCATTAAGCTCATCGACTTTGGCATCGCGACAGATCGAGGCTG 520
QY 384 TGCTGCCCTCAGACAAGTTCAGGACTACCTAGACCATGGCTCCAGAGCTCCTGGAGG 443
Db 521 GCAGCGAGTTCAGACATCTTTGGCACACCCGAGTTTGTGCCCCCGAGATCGTGA 580
QY 444 GCCAGGGGCTTTCCACAGACAGATCTGGGCCATCGGTGTGACAGCCCTTCAATATGC 503
Db 581 ATGAGCCACTTGGCTTGGAGGCTGACATGTGAGCAATTGGGCTCATCATCACTACATCTCC 640
QY 504 TGAGCGCGGAGTACCGGTGACAGCGAG 532
Db 641 TGAGCGGAGCTCCCAATTCCTGGGCGAG 669

RESULT 8

US-08-878-989-12

; Sequence 12, Application US/08878989

; Patent No. 5885803

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl G.

; APPLICANT: Lal, Preeti

; APPLICANT: Goli, Surya K.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN

; TITLE OF INVENTION: KINASES

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/878,989

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J J

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0321 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1282 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: PROSNOT06

; CLONE: 827431

US-08-878-989-12

Query Match 8.6%; Score 83.2; DB 2; Length 1282;

Best Local Similarity 48.2%; Pred. No. 1.2e-08;

Matches 235; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 59 GTGCTGGCGAATACAGAGCCCTCAAGGGCTCGCCACCCGCAACCCGCACTGCGCCAGCTGCAC 118
Db 372 GTGAGAACAGATCGCAGTGTCTCGTAGGATCAGTCAACCCCAACATCGCTCGCTCTGGAG 431
QY 119 GCAGCTTACCTCAGCCCCCGGACCTTCCACCTTACCTGGCCATGGAATGCTGAGCGGTGG 178
Db 432 GATGTCACAGAGAGCCCTTCCACCTTACCTGGCCATGGAATGCTGAGCGGTGGCGAG 491
QY 179 CTGCTCCCTGCTGCGCCGAGAGGGCTCTCTACTAGAATCCGAGTGAAGGACTACCTG 238
Db 492 CTGTTTACCGCATCATGGAGCGCGCTCTACACAGAGAAGGATGCCAGCCATCTGGTG 551
QY 239 TGGCAGATGTTGAGTGCCACCCAGTACCTGCACAACCCAGACATCCTGCACCTGGACCTG 298
Db 552 GGTGAGTCTTGGCGCGTCTCTACCTGCACAGCCTTGGGATCGTGACCCGGGACCTC 611
QY 299 AGGTCCGAGAACATGATCATCACCAATACCACTGCTCAAGGTCTGGAGCTTGGGCAAT 358
Db 612 AAGCCCGAACAACCTCTCTGTATGTCACGCCCTTTTGGAGACTCGAAGATCATGGTCTCTGAC 671
QY 359 GCACAGAGCTCAGCAGAGAGAGGTGCTGCCCTCAGACAGTTCAGGACTACCTAGAG 418
Db 672 TTTGGACTCTCCAAAATCCAGGCTGGGAAACATCTAGGCCACCCCTGTGGACCCCTGGA 731
QY 419 ACCATGGCTCCAGAGCTCTCTGGAGGGCCAGGGGCTGTTCACACAGACAGACATCTGGGCC 478
Db 732 TATGTGCCCCAGAGCTCTTGGAGAGAACCTTACGGNAGGCCGTAGATGTGTGGGCC 791
QY 479 ATCGGTGTGACGCCCTTATCATCTGAGCGCCGAGTACCCGGTGAAGCAGCGAGGTGCA 538
Db 792 CTGGCGTCTATCTCTACATCTCTGCTGTGTGGGTACCCCTTCTACGACGAGAGCGAC 851
QY 539 CGCGACCT 546
Db 852 CTGAGCT 859

RESULT 9

US-09-272-796-12

; Sequence 12, Application US/09272796

; Patent No. 8207148

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl G.

; APPLICANT: Lal, Preeti

; APPLICANT: Goli, Surya K.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN

; TITLE OF INVENTION: KINASES

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/272,796
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/878,989
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J J
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PR-0321 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 TELEX:
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1282 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: PROSNOT06
 CLONE: 827431
 US-09-272-796-12

Query Match 8.6%; Score 83.2; DB 3; Length 1282;
 Best Local Similarity 48.2%; Pred. NO. 1.2e-08;
 Matches 235; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 59 | GTGCTGCGCGAATAGAGGCGCTCAAGGGCTGCGCCACCGCACCTGGCCAGCTGCAC | 118 |
| DB | 372 | GTGGAGACGAGATCGAGTGTCTCGTAGGATCAGTACCCCAACATCGTCTGGAG | 431 |
| QY | 119 | GCAGCCTACCTAGCCCCCGGCACTGCTCATCTTTGGAGTGTGTCTTGGCCCGGAG | 178 |
| DB | 432 | GATGTCCACGAGAGCCCTTCCACCTCTACCTGGCCATGGAATCGTGGGTGGGAG | 491 |
| QY | 179 | GTGCTCCCTGCTGGCGAGAGGCGCTTCTACCTGAGATCGAGTGAAGGATCGGCAAT | 238 |
| DB | 492 | CIGTTGACCGATCATGAGCGGCTTCTACCTGAGATCGAGTGAAGGATCGGCAAT | 551 |
| QY | 239 | TGGCAGATGTTGAGTGCCACCGAGTACCTGCACACAGCAGCATCTGCACCTGGACCTG | 298 |
| DB | 552 | GTCAGGCTTGGCGCGCTTCTCTACCTGACAGCTGGGATCGTGCACCGGACCTC | 611 |
| QY | 299 | AGGTCGAGAACATGATCATCCGGAATACACCTGCTCAAGGTGTTGGAGCTGGGCAAT | 358 |
| DB | 612 | AAGCCGAAACCTCTGTATGCGACGCGCTTCTGAGGATCGAAGATCATGCTCTGAC | 671 |
| QY | 359 | GCACAGAGCCTCAGCCAGGAGAGTGTGCTCCCTCAGACAGTTCAGAGGACTACCTAGAG | 418 |
| DB | 672 | TTTGACTCTCCAAATCCAGGCTGGGAAATGCTAGGACCGCTTGTGGACCCCTGGA | 731 |
| QY | 419 | ACCATGCTCCAGAGCTCTTGGAGGCGGAGGCTGTTCACACAGACATCTGGCC | 478 |
| DB | 732 | TATGCGCCCGGAGAGCTCTTGGAGGAGAAACCTTACCGGAGGCGGCTGTAGTGTGGCC | 791 |
| QY | 479 | ATCGGTGACAGAGCCTTATCATGTGAGCGCCGAGTACCCGGTGGAGCAGCGAGGTGCA | 538 |
| DB | 792 | CTGGCGTTCATCTCTACATCTCTGTGTGGGTACCCCTTCTACGAGGAGGCGAC | 851 |
| QY | 539 | CGGACCT 546 | |
| DB | 852 | CCTGAGCT 859 | |

RESULT 10
 US-09-016-434-953
 : Sequence 953, Application US/09016434
 : Patent No. 6500938
 : GENERAL INFORMATION:
 : APPLICANT: Janice Au-Young
 : APPLICANT: Jeffrey J. Seilhamer
 : TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
 : TITLE OF INVENTION: PATHWAY GENE EXPRESSION
 : NUMBER OF SEQUENCES: 1490
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 : STREET: 3174 PORTER DRIVE
 : CITY: PALO ALTO
 : STATE: CALIFORNIA
 : COUNTRY: USA
 : ZIP: 94304
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/016,434
 : FILING DATE: HEREWITH
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER:
 : FILING DATE:
 : CLASSIFICATION:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Zeller, Karen J.
 : REGISTRATION NUMBER: 37,071
 : REFERENCE/DOCKET NUMBER: PA-0002 US
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (650) 855-0555
 : TELEFAX: (650) 845-4166
 : INFORMATION FOR SEQ ID NO: 953:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1282 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : IMMEDIATE SOURCE:
 : LIBRARY: PROSNOT06
 : CLONE: 827431
 : US-09-016-434-953

Query Match 8.6%; Score 83.2; DB 4; Length 1282;
 Best Local Similarity 48.2%; Pred. NO. 1.2e-08;
 Matches 235; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 59 | GTGCTGCGCGAATAGAGGCGCTCAAGGGCTGCGCCACCGCACCTGGCCAGCTGCAC | 118 |
| DB | 372 | GTGGAGACGAGATCGAGTGTCTCGTAGGATCAGTACCCCAACATCGTCTGGAG | 431 |
| QY | 119 | GCAGCCTACCTCAGCCCCCGGCACTGCTCATCTTTGGAGTGTGTCTTGGCCCGGAG | 178 |
| DB | 432 | GATGTCCACGAGAGCCCTTCCACCTCTACCTGGCCATGGAATCGTGGGTGGGAG | 491 |
| QY | 179 | GTGCTCCCTGCTGGCGAGAGGCGCTTCTACCTGAGATCGAGTGAAGGATCGGCAAT | 238 |
| DB | 492 | CIGTTGACCGATCATGAGCGGCTTCTCTACCTGACAGCTGGGATCGTGCACCGGACCTC | 551 |
| QY | 239 | TGGCAGATGTTGAGTGCCACCGAGTACCTGCACACAGCAGCATCTGCACCTGGACCTG | 298 |
| DB | 552 | GTCAGGCTTGGCGCGCTTCTCTACCTGACAGCTGGGATCGTGCACCGGACCTC | 611 |
| QY | 299 | AGGTCGAGAACATGATCATCCGGAATACACCTGCTCAAGGTGTTGGAGCTGGGCAAT | 358 |
| DB | 612 | AAGCCGAAACCTCTGTATGCGGACGCGCTTCTGAGGATCGAAGATCATGCTCTGAC | 671 |

QY 359 GCACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAG 418
Db 672 TTGGACTCTCCAAAATCCAGGCTGGAAACATCTAGGACCGCCCTGTGGGACCCCTGGA 731
QY 419 ACATGGCTCCAGAGCTCTGAGAGGCCAGGGGGGTGTTCACAGACAGACATCTGGGCC 478
Db 732 TATGTGGCCCCAGAGCTCTTGGAGCAGAAACCTACGGGAAGGCCGTAGATGTGTGGGCC 791
QY 479 ATCGGTGTACAGCCTTCATCATGTGTAGCGCGCGAGTACCCTGGTACGAGCAGGGTGCA 538
Db 792 CTGGGCTCATCTCTCATCTCTGTGTGGTACCCCTCTTCTACGACGAGCGAC 851
QY 539 CCGGACCT 546
Db 852 CCTGAGCT 859

RESULT 11
US-09-230-896C-5
; Sequence 5, Application US/09230896C
; Patent No. 6635479
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; APPLICANT: Sutcliffe, et al.
; TITLE OF INVENTION: Hypothalamus-Specific Polypeptides
; FILE REFERENCE: TSRI-548.1
; CURRENT APPLICATION NUMBER: US/09/230,896C
; CURRENT FILING DATE: 1999-02-02
; PRIOR APPLICATION NUMBER: 60/023,220
; PRIOR FILING DATE: 1996-08-02
; NUMBER OF SEQ. ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1458
; TYPE: DNA
; ORGANISM: ratus ratus
US-09-230-896C-5

Query Match 7.9%; Score 76.6; DB 4; Length 1458;
Best Local Similarity 50.1%; Pred. No. 2.9e-07;
Matches 246; Conservative 0; Mismatches 239; Indels 6; Gaps 2;

QY 59 GTGCTCGCGAATACGAGCCCTCAAGGGCTCGGCCACCCGACCTGGCCCGACGTGCAC 118
Db 221 GTGGAGATGAGATCGCAGTACTCGCAGAGATTAGCCACCCCAACATTTGGCTCTGGAG 280
QY 119 GCAGCTTACCTCAGCCCCCGGACCTGTGTCTCATCTTTGGAGCTGTGCTCTGGGCCGAG 178
Db 281 GACGTCCACGAGAGCCCTTCCCATCTCTACTTGGCCATGAGCTGGTAACAGGTGGTAA 340
QY 179 CTGCTCCCTGCTGCGCGAGAGGGCTCTCTACTCAGAACTCGAGGTGAGGACTACCTG 238
Db 341 CTGTTTGACCGAATCATGAGCGGGGCTCTACAGAGAGGATCGAGCCACCTTGTA 400
QY 239 TGGCAGATGTTGAGTSCCACCAGTACCTGCACACACAGACATCTGCACTGGAACCTG 298
Db 401 GGGCAGGTCTTGGTGTCT 460
QY 299 AGGTCCGAGAACATATC---ATCCCAATACACCTGTCTCAGGTGTGAGCTGGGCTGGC 355
Db 461 AAGCTGAAACCTCTCTATGCCACACCTTTTGGAGACTCCAGATCATGGTCTCTGAC 520
QY 356 AATGCACAGAGCTCAGCCAGGAGAGGTGTGCTGCTCCTCAGACAAAGTTCAAGGACTACT 415
Db 521 TTGGCTGTCCAAATTCAGCTGGCAACATGC---TAGGCAAGCTGTGGGACCCCA 577
QY 416 GAGACCATGGTCCAGAGCTCTTGAGGGCCAGGGGGGTGTTCACAGACAGACATCTGG 475
Db 578 GGATATGTGGCCCCAGAGCTCTCTGAGCAGAAACCTCAGCGGAAGGCCGTAGATGTGTGG 637
QY 476 GCATCGGTGTGACAGCTTCTATCATGTGTAGCGCCGAGTACCCGGTGGAGCAGGAGGT 535

Db 638 GCCTGGGTGTCTCTCTCTATCTCTATCTCTGCTGTGGGTACCCCTCTTCTATGATGAGAGC 697
QY 536 GCACGCGACCT 546
Db 698 GATCCTGAAC 708

RESULT 12
US-09-016-434-1147
; Sequence 1147, Application US/09016434
; Patent No. 650938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1147:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2169 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENEANK
; CLONE: g148262
US-09-016-434-1147

Query Match 7.3%; Score 71.2; DB 4; Length 2169;
Best Local Similarity 49.3%; Pred. No. 4.4e-06;
Matches 215; Conservative 0; Mismatches 218; Indels 3; Gaps 1;

QY 74 GAGGCCCTCAAGGGCTGCGCCACCCGACCTGGCCCGAGCTGCTCTGGGCGGAGCTGCTCCCTGCCTG 133
Db 253 GAGCTGCACCGAGACCTGCAGCACCCTCATCTGTCGCTTTTTCGCCACCATCTTGGAGAC 312
QY 134 CCGCGGACCTGCTGCTCATCTTGGAGCTGTGCTCTGGGCGGAGCTGCTCCCTGCCTG 193
Db 313 GCTGACAACTACATTTCTTGGAGCTCTGCAGCCGAAAGTCCCTGGCCACATCTCG 372
QY 194 GCGAGAGGCTCTCTACTCAGAAATCCGAGGTGAAGGACTACTGTGGCAGATGTTGAGT 253
Db 373 AAGGCCCGCACACCTCTGTTGGAGCCAGAAAGTGCCTACTACTCTGGCGGAGATCTCTTCT 432
QY 254 GCCACCCAGTACTCTGCACACAGCAGCATCTCCCTGCACCTGAGCTCCGAGACATG 313

Db 433 GGCCTCAAGTACTTGCACCGGGGCGCATCTTGCACCGGGACCTCAAGTTGGGAAATTT 492
QY 314 ATCATCCGGAATACAACTCTCAAGGTCTGGACCTGGGCAATGCACAGAGCTTCAGC 373
Db 493 TTCACTGAGAACTGGAAGTGAAGTGGGATTTGGGCTGGCA---GCCGGTTG 549
QY 374 CAGGAGAAGTGTCTCCCTCAGACAAGTTCAAGGACTACCTAGAGACCATGGTCCAGAG 433
Db 550 GAGCTCCGAGCAGAGGAAGAACCATCTGTGGCACCCTCACTATGTGGCTCCAGAA 609
QY 434 CTCCTGGAGGCGGAGGCGGCTTCCACAGACACATCTGGGCGCATCGGTGCAGAGCC 493
Db 610 GTGCTGCTGAGACAGGGCCGCGGCTGAGCGGATGTATGGTCACTGGGCTGTGTCATG 669
QY 494 TTCACTATGCTGAGCG 509
Db 670 TACACGCTGCTCTGG 685

RESULT 13
US-08-631-097-3
; Sequence 3, Application US/08631097
; Patent No. 5968816
; GENERAL INFORMATION:
; APPLICANT: Kimchi, Adi
; TITLE OF INVENTION: Tumor Suppressor Genes,
; TITLE OF INVENTION: Protein Enclosed Thereby, and Use of Said Genes and Protein
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wigman, Cohen, Leitner, & Myers, P.C.
; STREET: 900 17th Street, N.W., Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/631.097
; FILING DATE: 12-Apr-96
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/11598
; FILING DATE: 12-Oct-94
; ATTORNEY/AGENT INFORMATION:
; NAME: Cohen, Herbert
; REGISTRATION NUMBER: 25,109
; REFERENCE/DOCKET NUMBER: 0744.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)463-7700
; TELEFAX: (202)473-6915
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4935 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: No
; ANTI-SENSE: No
; FRAGMENT TYPE: No. 5968816 applicable
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; STRAIN: not applicable
; INDIVIDUAL ISOLATE: not applicable
; DEVELOPMENTAL STAGE: not applicable
; TISSUE TYPE: blood
; CELL TYPE: Leucocyte
; CELL LINE: HeLa

; ORGANELLE: not applicable
; IMMEDIATE SOURCE:
; LIBRARY: not applicable
; CLONE: not applicable
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: not applicable
; MAP POSITION: not applicable
; UNITS: not applicable
; FEATURE:
; NAME/KEY: Seq. ID. NO.: 3 is
; NAME/KEY: the sequence in claim 1(iii) as Figure 8 of the specification
; LOCATION: not available
; IDENTIFICATION METHOD: experiment-
; IDENTIFICATION METHOD: in specification
; OTHER INFORMATION: prevention of IFN-2
; OTHER INFORMATION: promoted cell death
; OTHER INFORMATION: not available
; PUBLICATION INFORMATION: not available
; US-08-631-097-3
Query Match 7.2%; Score 69.6; DB 2; Length 4935;
Best Local Similarity 53.3%; Pred. No. 1.2e-05;
Matches 147; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 64 GCGCGAATACGAGGCGCTCAAGGGCTGGCGCACCGGACCTGGCCACCTGGCCAGCTGCAGCAGC 123
Db 522 GCGGAGGTTCAGATCTCTGAGGAGATCCAGCACCCCAATGTCATCACCCTGCAGGAGT 581
QY 124 CTACTTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTCTTGGGCGGAGCTGCT 183
Db 582 CTATGAGAACAGACGAGCTCATCTGATCTTGAACCTCGTTCAGGTGGGAGCTGTT 641
QY 184 CCCCTGCTGGCGGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGTGGCA 243
Db 642 TGACTTCTTAGCTGAAGAGATCTTTAACTGAAGAGGAGCAACTGAATTTCAACA 701
QY 244 GATGTTGAGTGCCACCCAGTACCTGCACAAACAGACATCTCTGCACCTGGACCTGAGTC 303
Db 702 AATTCCTAATGTTTACTACCTGCACTCCCTTCAATCGCCCACTTTGATCTTAAGCC 761
QY 304 CGAGACATGATCATCACCAGATACACCTGCTCAA 339
Db 762 TGAGAACATAATGCTTTGGATAGAAATGTCCTCCAA 797
RESULT 14
US-08-810-712-9
; Sequence 9, Application US/08810712G
; Patent No. 6160106
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. LTD
; TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and
; TITLE OF INVENTION: Use of said Genes and Proteins
; FILE REFERENCE: sequencelist
; CURRENT APPLICATION NUMBER: US/08/810,712G
; CURRENT FILING DATE: 1997-03-03
; EARLIER APPLICATION NUMBER: PCT/US94/11598
; EARLIER FILING DATE: 1994-10-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 9
; LENGTH: 5886
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (337)..(4605)
; US-08-810-712-9
Query Match 7.2%; Score 69.6; DB 3; Length 5886;
Best Local Similarity 53.3%; Pred. No. 1.3e-05;
Matches 147; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 64 GCGCGAATACGAGGCGCTCAAGGGCTGGCGCACCGGACCTGGCCAGCTGCAGCAGC 123

Db 522 GCGGAGTCAAGATCCCTGAGGAGATCCAGACCCCAATGTCATCACCTGACAGGT 581
Qy 124 CTACCTCAGCCCGCGACCTGGTGTCTATCTTGGAGCTGTGTCTGGGCGCCGAGCTGT 183
Db 582 CTATGAGAACAGACGAGCTGTCCTGATCTTGGAACTCGTTGACGCTGGCGAGCTGT 641
Qy 184 CCCTGCTGCGCAGAGGGCTCTCTACTCAGATCCGAGTGAAGGACTACCTGTGGCA 243
Db 642 TGACTTCTTAGCTGAAAAGGAATCTTTAACTGAAGAGCAACTGAAATTTCTCAAACA 701
Qy 244 GATGTTGAGTGCCACCCAGTACCTGCACAAACCCAGCACATCTGCACCTGGACCTGAGGTC 303
Db 702 AATCTTAAATGGTGTCTTACTTACTGCACTCCCTTCAATCGCCACCTTTGATCTTAAGCC 761
Qy 304 CGAAGACATGATCATACCCGAAATACACCTGCTCAA 339
Db 762 TGAGAACATATGCTTTTGGATAGAAATGTCCCCAA 797

RESULT 15

US-09-579-664B-5
; Sequence 5, Application US/09579664B
; Patent No. 6514719
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Bird, Timothy A.
; APPLICANT: Virca, G. Duke
; APPLICANT: Martin, Unja
; APPLICANT: Anderson, Dirk M.
; TITLE OF INVENTION: NOVEL MURINE AND HUMAN KINASES
; FILE REFERENCE: 2923-A
; CURRENT APPLICATION NUMBER: US/09/579,664B
; CURRENT FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 3228
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-579-664B-5

Query Match 7.1%; Score 69; DB 4; Length 3228;
Best Local Similarity 47.6%; Pred. No. 1.4e-05;
Matches 204; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

Qy 60 TGCTGCGGGAATACAGGCCCTCAAGGGCCCTGCGCCACCCGACCTGGCCCGAGCTGCACG 119
Db 296 TCCTGACTGAGATTGAGATCTCTAAGGGCATTCGGCACCCCATATCGTGACGCTGAAG 355
Qy 120 CAGCCTACCTCAGCCCGCGCACCTGGTGCTCATCTTGGAGCTGTGTCTGGGCGCCGAGC 179
Db 356 ACTTCCAGTGGGACAAATGACAATATCTACCTCATCATGGAGTTCTGTGACGGGGTGACC 415
Qy 180 TGCTCCCTGCTGCGCCGAGGGCCCTCTACTCAGATCCGAGGTGAAGACTACCTGT 239
Db 416 TGTCTCGCTTCATTCAACCCGAGGATCTGCTGAGAGGTGGCCCGTGTTCATGC 475
Qy 240 GGCAGATGTTGAGTGCCACCCAGTACCTGCACAAACCCAGCACATCTGCACCTGGACCTGA 299
Db 476 AGCAGTTGGCTAGTGCCTCGAGTTCTCTGATGAGCAACATCTCTCACTTGGATCTGA 535
Qy 300 GGTCCGAGAACATGATCATACCGAATACACCTGCTCAAGTGTGTGGACCTGGGCAATG 359
Db 536 AACCCGAGAACATCTCTGTGAGCTCTTTGGAGAAAGCCCACTGAAACTGGCAGACTTG 595
Qy 360 CACAGAGCCTCAGCCAGGAGAGGTGTGCTGCCCTCAGACAAAGTTCAGAGACTACCTAGAGA 419
Db 596 GCTTTGCCAGCACATGTCCTCCCGTGGAGAGAAACAGTGTCTCCGTGGCTCCCGGCTCT 655
Qy 420 CCATGGCTCCAGAGTCTCTGAGAGGCGCACGGGGCTGTTCACAGACAGACATCTGGGCCA 479
Db 656 ATATGGCTCCTGAGATGCTGTGTCGGCGGCGAGTATGATGCGGCTGTGGACCTCTGGTCTG 715

Qy 480 TCGGTGTGA 488
Db 716 TGGGGTGA 724
Search completed: September 19, 2004, 22:12:23
Job time : 93.2015 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 19:28:00 ; Search time 502.587 Seconds
(without alignments)
9751.690 Million cell updates/sec

Title: US-10-077-130-4 COPY 23150 24120

Perfect score: 971
Sequence: 1 cagcggcgggcgctgaccg.....ggccgctaaaaagtctaga 971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications NA:**

- 1: /cgn2_5/prodata/2/pubpna/us07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/2/pubpna/us06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/2/pubpna/us06_PUBCOMB.seq.*
- 5: /cgn2_6/prodata/2/pubpna/us07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/prodata/2/pubpna/us08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/2/pubpna/us08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/2/pubpna/us09A_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/2/pubpna/us09B_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/2/pubpna/us09C_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/2/pubpna/us09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/2/pubpna/us09_PUB.seq.*
- 14: /cgn2_6/prodata/2/pubpna/us10A_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/2/pubpna/us10B_PUBCOMB.seq.*
- 16: /cgn2_6/prodata/2/pubpna/us10C_PUBCOMB.seq.*
- 17: /cgn2_6/prodata/2/pubpna/us10_NEW_PUB.seq.*
- 18: /cgn2_6/prodata/2/pubpna/us60_NEW_PUB.seq.*
- 19: /cgn2_6/prodata/2/pubpna/us60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | | | DB | ID | Description |
|------------|-------|-------|--------|----|---------------------|--------------------|
| | Score | Match | Length | | | |
| 1 | 971 | 100.0 | 8106 | 14 | US-10-077-130-1 | Sequence 1, Appli |
| 2 | 971 | 100.0 | 24120 | 14 | US-10-077-130-4 | Sequence 4, Appli |
| 3 | 946.4 | 97.5 | 4475 | 15 | US-10-307-019-2 | Sequence 2, Appli |
| 4 | 926.8 | 95.4 | 5207 | 9 | US-09-858-664A-1 | Sequence 1, Appli |
| 5 | 926.8 | 95.4 | 5207 | 13 | US-10-697-263-1 | Sequence 1, Appli |
| 6 | 926.8 | 95.4 | 5207 | 13 | US-10-697-263-3 | Sequence 3, Appli |
| 7 | 911.4 | 93.9 | 3225 | 17 | US-10-311-034-46 | Sequence 46, Appli |
| 8 | 911.4 | 93.9 | 5454 | 13 | US-10-415-011-44 | Sequence 44, Appli |
| 9 | 906.4 | 93.3 | 1253 | 13 | US-10-435-114-26278 | Sequence 26278, A |
| 10 | 906.4 | 93.3 | 4336 | 13 | US-10-182-243-14 | Sequence 14, Appli |
| 11 | 905.4 | 93.2 | 5007 | 15 | US-10-307-019-3 | Sequence 3, Appli |
| 12 | 905.4 | 93.2 | 7828 | 15 | US-10-307-019-5 | Sequence 5, Appli |
| 13 | 829 | 85.4 | 7893 | 14 | US-10-077-130-3 | Sequence 3, Appli |
| 14 | 829 | 85.4 | 23907 | 14 | US-10-077-130-6 | Sequence 6, Appli |

| | | | | | | |
|----|-------|------|-------|----|---------------------|---------------------|
| 15 | 223.4 | 23.0 | 225 | 11 | US-09-864-408A-2403 | Sequence 2403, Appl |
| 16 | 204.4 | 21.1 | 7789 | 13 | US-10-333-334-38 | Sequence 38, Appl |
| 17 | 204.4 | 21.1 | 9807 | 16 | US-10-379-381-1 | Sequence 1, Appl |
| 18 | 202.8 | 20.9 | 9698 | 16 | US-10-110-130-37 | Sequence 37, Appl |
| 19 | 202.8 | 20.9 | 10122 | 16 | US-10-210-130-35 | Sequence 35, Appl |
| 20 | 191.8 | 19.8 | 1627 | 13 | US-10-425-114-26265 | Sequence 26265, A |
| 21 | 135.6 | 14.0 | 9930 | 16 | US-10-210-130-33 | Sequence 33, Appl |
| 22 | 113.8 | 11.7 | 2254 | 15 | US-10-059-585-11 | Sequence 11, Appl |
| 23 | 112.2 | 11.6 | 2105 | 17 | US-10-468-406-17 | Sequence 17, Appl |
| 24 | 112.2 | 11.6 | 2132 | 16 | US-10-463-445-118 | Sequence 118, App |
| 25 | 89.2 | 9.2 | 1788 | 9 | US-09-797-039-9 | Sequence 9, Appl |
| 26 | 89.2 | 9.2 | 1788 | 15 | US-10-170-789-9 | Sequence 9, Appl |
| 27 | 89.2 | 9.2 | 1791 | 13 | US-10-423-543-26 | Sequence 26, Appl |
| 28 | 89.2 | 9.2 | 1791 | 17 | US-10-380-236A-1 | Sequence 1, Appl |
| 29 | 89.2 | 9.2 | 2046 | 9 | US-09-797-039-7 | Sequence 7, Appl |
| 30 | 89.2 | 9.2 | 2046 | 13 | US-10-423-543-24 | Sequence 24, Appl |
| 31 | 89.2 | 9.2 | 2046 | 15 | US-10-170-789-7 | Sequence 7, Appl |
| 32 | 89.2 | 9.2 | 3199 | 13 | US-10-112-944-49 | Sequence 49, Appl |
| 33 | 89 | 9.2 | 860 | 16 | US-10-210-130-39 | Sequence 39, Appl |
| 34 | 88 | 9.1 | 1289 | 17 | US-10-437-963-76904 | Sequence 76904, A |
| 35 | 87.6 | 9.1 | 1300 | 13 | US-10-425-114-30892 | Sequence 30892, A |
| 36 | 84.6 | 8.7 | 1788 | 10 | US-09-849-138-7 | Sequence 7, Appl |
| 37 | 84.6 | 8.7 | 1971 | 10 | US-09-849-138-31 | Sequence 1, Appl |
| 38 | 84.6 | 8.7 | 2558 | 10 | US-09-849-138-5 | Sequence 5, Appl |
| 39 | 84.6 | 8.7 | 2558 | 10 | US-09-849-138-5 | Sequence 5, Appl |
| 40 | 83.2 | 8.6 | 1282 | 10 | US-09-765-970-12 | Sequence 12, Appl |
| 41 | 83.2 | 8.6 | 1282 | 16 | US-10-302-720-553 | Sequence 953, App |
| 42 | 83.2 | 8.6 | 1582 | 13 | US-10-425-114-26649 | Sequence 26649, A |
| 43 | 83.2 | 8.6 | 1584 | 13 | US-10-302-172-205 | Sequence 205, App |
| 44 | 82.8 | 8.5 | 2973 | 15 | US-10-311-034-32 | Sequence 32, Appl |
| 45 | 82.6 | 8.5 | 1461 | 15 | US-10-156-761-3527 | Sequence 3527, Ap |

ALIGNMENTS

```

RESULT 1
US-10-077-130-1
; Sequence 1, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 53079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1RCP1(W)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 8106
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(7964)
; NAME/KEY: 3'UTR
; LOCATION: (7965)...(8106)
US-10-077-130-1

```

```
Query Match      100.0%; Score 971; DB 14; Length 8106;
Best Local Similarity 100.0%; Pred. No. 1.8e-220;
Matches 971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

| Qy | 1 | CAGCGGCGGCGCTGGCGCGCAAGATCATCCCCCTACCCACCCCAAGACAAGACAGCAGT | 60 |
|----|------|---|------|
| Dy | 7136 | CAGCGGCGGCGCTGGCGCGCAAGATCATCCCCCTACCCACCCCAAGACAAGACAGCAGT | 7195 |

Db 7136 CAGCGGCGGGCGCTGGCCGCCAAGATCATCCCCCTACCACCCCAAGGACAAGCAGCAGT 7195

61 GCTGGCGAATACAGAGCCCTCAAGGGCTTGGCCACCCGACCTGGCCAGCTGCACGC 120
Db GCTGGCGAATACAGAGCCCTCAAGGGCTTGGCCACCCGACCTGGCCAGCTGCACGC 7355
121 AGCTACTCTAGCCCCCGGCACTGGTCTCATCTTGGAGTGTGCTCTGGGCCCCAGCT 180
Db AGCTACTCTAGCCCCCGGCACTGGTCTCATCTTGGAGTGTGCTCTGGGCCCCAGCT 7315
181 GCTCCCTCTGGCCGAGAGGGCTCTCTCTCAGAAATCCAGTGAAGGACTACCTGTG 240
Db GCTCCCTCTGGCCGAGAGGGCTCTCTCTCAGAAATCCAGTGAAGGACTACCTGTG 7375
241 GCAGATGTTGAGTGCACCCAGTACCTGCACAACCCAGCACATCTCTGCACTGGACCTGAG 300
Db GCAGATGTTGAGTGCACCCAGTACCTGCACAACCCAGCACATCTCTGCACTGGACCTGAG 7435
301 GTCGAGAACATGATCATACCGAATACAACTCTCAAGTCTGAGTCTGGGCAATGC 360
Db GTCGAGAACATGATCATACCGAATACAACTCTCAAGTCTGAGTCTGGGCAATGC 7495
361 ACAGAGCTCTAGCCAGAGAGAGTGTCTGCTCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420
Db ACAGAGCTCTAGCCAGAGAGAGTGTCTGCTCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 7555
421 CATGGCTCCAGAGCTCTGGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCAAT 480
Db CATGGCTCCAGAGCTCTGGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCAAT 7615
481 CGGTGTGACAGCTTCACTCATCTGAGCGCCGAGTACCCGGTGAAGAGGCTGCACG 540
Db CGGTGTGACAGCTTCACTCATCTGAGCGCCGAGTACCCGGTGAAGAGGCTGCACG 7675
541 CGACTCTGAGAGAGCTGCGCAAGGGGCTGTGTCGGCTGAGCGGCTGCTACGCGGGCT 600
Db CGACTCTGAGAGAGCTGCGCAAGGGGCTGTGTCGGCTGAGCGGCTGCTACGCGGGCT 7735
601 GTCCGGGGCGCGTGGCTTCTCTGCGCAGACACTGTGTGCGCCAGCCCTGGGGCGGGCC 660
Db GTCCGGGGCGCGTGGCTTCTCTGCGCAGACACTGTGTGCGCCAGCCCTGGGGCGGGCC 7795
661 CTGCGCTCCAGCTGCTCTGAGTCCCGTGTCTAAACAGAGAGGGCCCGCTGTTCGCG 720
Db CTGCGCTCCAGCTGCTCTGAGTCCCGTGTCTAAACAGAGAGGGCCCGCTGTTCGCG 7855
721 GCCCGCGCGCTGACCTTCCCTACCGCGCGCTGGCGCTTCTGCTGCGCAATGCGAGAA 780
Db GCCCGCGCGCTGACCTTCCCTACCGCGCGCTGGCGCTTCTGCTGCGCAATGCGAGAA 7915
781 GAGAGCGCGCTGTCTATACAGAGAGCAAACTGGCCCGAGTGTGCTGAGGGTGCAGCGCG 840
Db GAGAGCGCGCTGTCTATACAGAGAGCAAACTGGCCCGAGTGTGCTGAGGGTGCAGCGCG 7975
841 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGCAGACGGCCCAATAAAGCCGACG 900
Db GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGCAGACGGCCCAATAAAGCCGACG 8035
901 CGGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 960
Db CGGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 8035
961 AAAAGTCTAGA 971
Db AAAAGTCTAGA 8106

RESULT 2

US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 53079 and 12599, Protein Kinase Family

; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
US-10-077-130-4

Query Match 100.0%; Score 971; DB 14; Length 24120;
Best Local Similarity 100.0%; Pred. No. 2.3e-220;
Matches 971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCGGGCGGGCGCTGGCGCCCAAGATCATCCCTTACCACCCCAAGGACAAGACGAGT 60
Db 23150 CAGCGGGCGGGCGCTGGCGCCCAAGATCATCCCTTACCACCCCAAGGACAAGACGAGT 23209
QY 61 GCTGCGGAATACGAGGGCCCTCAAGGGCTGTGGCCACCGCACCTGGCCAGCTGCACG 120
Db 23210 GCTGCGGAATACGAGGGCCCTCAAGGGCTGTGGCCACCGCACCTGGCCAGCTGCACG 23269
QY 121 AGCTACTCTAGCCCCCGGCACTGGTGTCTCATCTTGGAGTGTGCTTGGGCCCGAGCT 180
Db 23270 AGCTACTCTAGCCCCCGGCACTGGTGTCTCATCTTGGAGTGTGCTTGGGCCCGAGCT 23329
QY 181 GCTCCCTCTGGCCGAGAGGGCTCTCTCTCAGAAATCCGAGTGAAGGACTACCTGTG 240
Db 23330 GCTCCCTCTGGCCGAGAGGGCTCTCTCTCAGAAATCCGAGTGAAGGACTACCTGTG 23389
QY 241 GCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGACACATCTCTGCACTGGACCTGAG 300
Db 23390 GCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGACACATCTCTGCACTGGACCTGAG 23449
QY 301 GTCGAGAACATGATCATACCGAATACAACTCTCAAGTCTGAGTCTGGACCTGGCAATGC 360
Db 23450 GTCGAGAACATGATCATACCGAATACAACTCTCAAGTCTGAGTCTGGACCTGGCAATGC 23509
QY 361 ACAGAGCTCTAGCCAGGAGAGGCTGCTGCTCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420
Db 23510 ACAGAGCTCTAGCCAGGAGAGGCTGCTGCTCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 23569
QY 421 CATGGCTCCAGAGCTCTCTGGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCAAT 480
Db 23570 CATGGCTCCAGAGCTCTCTGGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCAAT 23629
QY 481 CGGTGTGACAGCTTCACTCATCTGAGCGCCGAGTACCCGGTGAAGGACTACCTAGAGAC 540
Db 23630 CGGTGTGACAGCTTCACTCATCTGAGCGCCGAGTACCCGGTGAAGGACTACCTAGAGAC 23689
QY 541 CGACTCTGAGAGAGCTGCGCAAGGGGCTGTGCTGCGCTGAGCGCTGCTACCGGGGCT 600
Db 23690 CGACTCTGAGAGAGCTGCGCAAGGGGCTGTGCTGCGCTGAGCGCTGCTACCGGGGCT 23749
QY 601 GTCCGGGGCGCGCTGGGCTTCTCTGCGCAGACACTGTGTGCGCCCGAGCCCTGGGGCGGGCC 660
Db 23750 GTCCGGGGCGCGCTGGGCTTCTCTGCGCAGACACTGTGTGCGCCCGAGCCCTGGGGCGGGCC 23809
QY 661 CTGCGCTCCAGCTGCTCTGAGTCCCGTGTCTAACAGAGAGGGCCCGCTGTTCGCG 720
Db 23810 CTGCGCTCCAGCTGCTCTGAGTCCCGTGTCTAACAGAGAGGGCCCGCTGTTCGCG 23869

QY 721 GCCCGCGCGCTGACCTTCCCTACCGCGCGGTGCGCGTCTTCTGTCGCAATCGAGAA 780
Db 23870 GCCCGCGCGGTGACCTTCCCTACCGCGCGGTGCGCGTCTTCTGTCGCAATCGAGAA 23929
QY 781 GAGAGCGCGGTGCTGTACAGAGGACAACTGCGCCCGAGTGGCTGAGGGTGCCTCCG 840
Db 23930 GAGAGCGCGGTGCTGTACAGAGGACAACTGCGCCCGAGTGGCTGAGGGTGCCTCCG 23989
QY 841 GCCACACCTTGGTCTCCCGCTGCGGGGTGCTGACAGCGCGCAATAAAAAACGCCCAGC 900
Db 23990 GCCACACCTTGGTCTCCCGCTGCGGGGTGCTGACAGCGCGCAATAAAAAACGCCCAGC 24049
QY 901 CGGGCGAGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 960
Db 24050 CGGGCGAGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 24109
QY 961 AAAAGTCTAGA 971
Db 24110 AAAAGTCTAGA 24120

RESULT 3

US-10-307-019-2
; Sequence 2, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenlin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SCIOS, INC.
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS 021DV1
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 4175
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(4053)
US-10-307-019-2

Query Match 97.5%; Score 946.4; DB 15; Length 4175;
Best Local Similarity 99.9%; Pred. No. 1.1e-214;
Matches 947; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGCGGGGCTGGCGCGCAAGATCATCCCTACACCCCAAGGCAAGACAGCAGT 60
Db 3228 CAGCGGCGGGGCTGGCGCGCAAGATCATCCCTACACCCCAAGGCAAGACAGCAGT 3287
QY 61 GTGCGCAATACGAGGCGCTCAAGGGCGCTGCGCCACCGACCTGGCGCCAGCTGCAGC 120
Db 3288 GTGCGCAATACGAGGCGCTCAAGGGCGCTGCGCCACCGACCTGGCGCCAGCTGCAGC 3347
QY 121 AGCTACCTCAGCCCCCGGCACTGGTGTCTCATCTTGTAGCTGTGCTCTGGGCCCGAGCT 180
Db 3348 AGCTACCTCAGCCCCCGGCACTGGTGTCTCATCTTGTAGCTGTGCTCTGGGCCCGAGCT 3407
QY 181 GTCCCTCGCTGGCGAGGCGCTCTACTCAGATCCGAGGTGAAGGACTACTGTG 240
Db 3408 GTCCCTCGCTGGCGAGGCGCTCTACTCAGATCCGAGGTGAAGGACTACTGTG 3467
QY 241 GCAGATGTTGAGTGCCACCCAGTACCTGCACCAACAGCAGCATCTGCACTGGACCTGAG 300
Db 3468 GCAGATGTTGAGTGCCACCCAGTACCTGCACCAACAGCAGCATCTGCACTGGACCTGAG 3527
QY 301 GTCCGAGAACATGATCATCACCAGATACAACTGCTCAAGGTCTGGGACCTGGCAATGC 360

Db 3528 GTCCGAGAACATGATCATCACCAGATACAACTGCTCAAGGTCTGGGCAATGC 3587
QY 361 ACAGAGCCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTTACCTAGAGAC 420
Db 3588 ACAGAGCCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTTACCTAGAGAC 3647
QY 421 CATGGTCCAGAGCTCCTGAGAGGCGCAGGGGGGTGTTCCACAGACAGACATCTGGGCCAT 480
Db 3648 CATGGTCCAGAGCTCCTGAGAGGCGCAGGGGGGTGTTCCACAGACAGACATCTGGGCCAT 3707
QY 481 CGGTGTGACAGCCTTTCATATGCTGAGCGCGGAGTACCCCGTGAGGAGCGAGGTTGACG 540
Db 3708 CGGTGTGACAGCCTTTCATATGCTGAGCGCGGAGTACCCCGTGAGGAGCGAGGTTGACG 3767
QY 541 CGACCTGACAGAGGAGCTGCGCAAGGGGTGTTCCGCTGAGCGCTGCTACGCGGGCT 600
Db 3768 CGACCTGACAGAGGAGCTGCGCAAGGGGTGTTCCGCTGAGCGCTGCTACGCGGGCT 3827
QY 601 GTCCGGGGGCGCGTGGCGCTTCTGCGCAGCACTCTGCGCCAGCCCTGGGGCGCGCC 660
Db 3828 GTCCGGGGGCGCGTGGCGCTTCTGCGCAGCACTCTGCGCCAGCCCTGGGGCGCGCC 3887
QY 661 CTGCGGCTCCAGCTGCTGCAAGGCGCGCTGCGCTTAAACAGAGGAGGCGCGCTGTTGCGG 720
Db 3888 CTGCGGCTCCAGCTGCTGCAAGGCGCGCTGCGCTTAAACAGAGGAGGCGCGCTGTTGCGG 3947
QY 721 GCCCGCGCGCTGACCTTCCCTACCGCGCGGTGCGCGTCTTCTGCGCAATCGCGAGAA 780
Db 3948 GCCCGCGCGCTGACCTTCCCTACCGCGCGGTGCGCGTCTTCTGCGCAATCGCGAGAA 4007
QY 781 GAGACGCGCGCTGCTGTACAGAGGAGCACAACCTGCGCCAGTGCCTGAGGGTGCCTCCG 840
Db 4008 GAGACGCGCGCTGCTGTACAGAGGAGCACAACCTGCGCCAGTGCCTGAGGGTGCCTCCG 4067
QY 841 GCCACACCTTGGTCTCCCGCTGCGGGGTGCTGACAGCGCGCAATAAAAAACGCCCAGC 900
Db 4068 GCCACACCTTGGTCTCCCGCTGCGGGGTGCTGACAGCGCGCAATAAAAAACGCCCAGC 4127
QY 901 CGGGCGAGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 948
Db 4128 CGGGCGAGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 4175

RESULT 4

US-09-858-664A-1
; Sequence 1, Application US/09858664A
; Patent No. US20020072491A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL00927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-858-664A-1

Query Match 95.4%; Score 926.8; DB 9; Length 5207;
Best Local Similarity 99.8%; Pred. No. 5.1e-210;
Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGCGGCGGGGCTGGCGCGCAAGATCATCCCTACACCCCAAGGCAAGACAGCAGT 60
Db 4278 CAGCGGCGGGGCTGGCGCGCAAGATCATCCCTACACCCCAAGGCAAGACAGCAGT 4337

Db 5058 GAGACGCGCGTCTGTATCAAGAGGCACAACTGGCCCGAGGTGGCTGAGGGTGCCTCCG 5117
Qy 841 GCCACACCTTGGTCTCCCGCTGGGGTGGCTGAGAGCGCGCAATAAAAAACGCCAGC 900
Db 5118 GCCACACCTTGGTCTCCCGCTGGGGTGGCTGAGAGCGCGCAATAAAAAACGCCAGC 5177
Qy 901 CGGCGAGAAAAAATAAAAAAATAAAAAA 930
Db 5178 CGGCGAGAAAAAATAAAAAAATAAAAAA 5207

RESULT 6

US-10-697-263-3
; Sequence 3, Application US/10697263
; Publication No. US20040063142A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00927-CIP-DIV2
; CURRENT APPLICATION NUMBER: US/10697,263
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: 10/274,978
; PRIOR FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-697-263-3

Query Match 95.4%; Score 926.8; DB 13; Length 5207;
Best Local Similarity 99.8%; Pred. No. 5.1e-210; Mismatches 2; Indels 0; Gaps 0;
Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CAGCGGGGGGGCGCTGGCGCCCAAGATCATCCCTTACCACCCCAAGGACAGACGACGT 60
Db 4278 CAGCGGGGGGGCGCTGGCGCCCAAGATCATCCCTTACCACCCCAAGGACAGACGACGT 4337
Qy 61 GCTGGCGCAATACGAGGGCCCTCAAGGGCTGGCGCCCAAGGACAGACGACGTGCAGC 120
Db 4338 GCTGGCGCAATACGAGGGCCCTCAAGGGCTGGCGCCCAAGGACAGACGACGTGCAGC 4397
Qy 121 AGCCTACCTCAGCGCCCGGCGCTGCTCATCTTTGAGGCTGCTCTGGGCGCGAGCT 180
Db 4398 AGCCTACCTCAGCGCCCGGCGCTGCTCATCTTTGAGGCTGCTCTGGGCGCGAGCT 4457
Qy 181 GTCCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGAACTCGAGAGTGAAGGACTACCTGTG 240
Db 4458 GTCCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGAACTCGAGAGTGAAGGACTACCTGTG 4517
Qy 241 GCAGATGTTGAGTGCACCCAGTACTGCACAAACAGACATCTCGACCTGGACCTGAG 300
Db 4518 GCAGATGTTGAGTGCACCCAGTACTGCACAAACAGACATCTCGACCTGGACCTGAG 4577
Qy 301 GTCGAGAAATGATCATCACCAGATACAACTGCTCAAGGTCGAGGCTGGGCAATGC 360
Db 4578 GTCGAGAAATGATCATCACCAGATACAACTGCTCAAGGTCGAGGCTGGGCAATGC 4637
Qy 361 ACAGAGCTCAGCAGGAGAGGTCGCTGCTCCAGAGTTCAGAGTACCTAGAGAC 420
Db 4638 ACAGAGCTCAGCAGGAGAGGTCGCTGCTCCAGAGTTCAGAGTACCTAGAGAC 4697
Qy 421 CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480
Db 4698 CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 4757

Qy 481 CGGTGTGACAGACCTTCATCATCTGAGCGCCGAGTACCCGGTGAAGCAGGAGGTGCACG 540
Db 4758 CGGTGTGACAGACCTTCATCATCTGAGCGCCGAGTACCCGGTGAAGCAGGAGGTGCACG 4817
Qy 541 CGACCTGCAGAGAGGACTGCGCAAGGGGCTGGTCCGGCTGAGCGCTGAGCGGGCT 600
Db 4818 CGACCTGCAGAGAGGACTGCGCAAGGGGCTGGTCCGGCTGAGCGCTGAGCGGGCT 4877
Qy 601 GTCGGGGGGCGCGCTGGCTTCTTCTGCGCAGCACTCTGTGCGCCACGCTTGGGGCGCGGC 660
Db 4878 GTCGGGGGGCGCGCTGGCTTCTTCTGCGCAGCACTCTGTGCGCCACGCTTGGGGCGCGGC 4937
Qy 661 CTGCGGCTCAGCTGCTGAGTCCGCTGGCTTAACAGAGAGGGGCGCGCTGTTTCGGG 720
Db 4938 CTGCGGCTCAGCTGCTGAGTCCGCTGGCTTAACAGAGAGGGGCGCGCTGTTTCGGG 4997
Qy 721 GCCCGCGCGCTGACCTTCCCTACCGCGGGCTGCGCTCTTCTGCGCAATCGCGAGAA 780
Db 4998 GCCCGCGCGCTGACCTTCCCTACCGCGGGCTGCGCTCTTCTGCGCAATCGCGAGAA 5057
Qy 781 GAGACGCGCTGCTGTATCAAGAGGACAACTTGGCGCCAGGTGCGCTGAGGGTGCCTCCG 840
Db 5058 GAGACGCGCTGCTGTATCAAGAGGACAACTTGGCGCCAGGTGCGCTGAGGGTGCCTCCG 5117
Qy 841 GCCACACCTTGGTCTCCCGCTGGGGTGGCTGCGAGACGCGCAATAAAAAACGCCAGC 900
Db 5118 GCCACACCTTGGTCTCCCGCTGGGGTGGCTGCGAGACGCGCAATAAAAAACGCCAGC 5177
Qy 901 CGGCGAGAAAAAATAAAAAAATAAAAAA 930
Db 5178 CGGCGAGAAAAAATAAAAAAATAAAAAA 5207

RESULT 7

US-10-311-034-46
; Sequence 46, Application US/10311034
; Publication No. US20040023242A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
; APPLICANT: BANDMAN, Olga
; APPLICANT: BOROMSKY, Mark L.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: LU, Yan
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: YAO, Monique G.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: GREENWALD, Sara R.
; APPLICANT: RANKUMAR, Javalaxmi
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: KEARNEY, Liam
; APPLICANT: BURFORD, Neil
; APPLICANT: NGUYEN, Dannel B.
; APPLICANT: TANG, Y. Tom
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: HE, Ann
; APPLICANT: THORNTON, Michael
; APPLICANT: HAPALIA, April
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: LO, Terence P.
; APPLICANT: KHAH, Farrah A.
; APPLICANT: RECIPON, Shirley A.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: POLICKY, Jennifer L.
; APPLICANT: DING, Li
; APPLICANT: GREYER, Megan
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: BATRA, Sajeev

APPLICANT: ISON, Craig H.
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0125 PCT
; CURRENT APPLICATION NUMBER: US/10/311,034
; PRIOR FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 60/212,073; 60/213,467; 60/215,651; 60/216,605; 60/218,372;
; 60/228,056
; PRIOR FILING DATE: 2000-06-15; 2000-06-23; 2000-06-30; 2000-07-07; 2000-07-13; 2000-07-25
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PERL Program
; SEQ ID NO 46
; LENGTH: 3225
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040023242A1 3016969CB1
US-10-311-034-46

Query Match 93.9%; Score 911.4; DB 17; Length 3225;
Best Local Similarity 99.9%; Pred. No. 2.1e-206;
Matches 912; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGCCAGATCATCCCTTACACCCCAAGGACAGACAGT 60
DB 2313 CAGCGGGGGGGCTGGCGCCAGATCATCCCTTACACCCCAAGGACAGACAGT 2372
QY 61 GCTGGCGGAATACGAGGGCCCTCAAGGCGCTGGCGCCACCCGACCTGGCCACCTGTCACGC 120
DB 2373 GCTGGCGGAATACGAGGGCCCTCAAGGCGCTGGCGCCACCCGACCTGGCCACCTGTCACGC 2432
QY 121 AGCTTACTCAGCCCCCGGACCTGGTGTCTCATCTTGGAGTGTGCTTGGGCGGAGCT 180
DB 2433 AGCTTACTCAGCCCCCGGACCTGGTGTCTCATCTTGGAGTGTGCTTGGGCGGAGCT 2492
QY 181 GCTCCCTGCTGGCGGAGAGGGCTCTTACTCAGAACTCCAGGCTGAAGGACTACCTGTG 240
DB 2493 GCTCCCTGCTGGCGGAGAGGGCTCTTACTCAGAACTCCAGGCTGAAGGACTACCTGTG 2552
QY 241 CGAGATGTTAGTGCACCCAGTACTGCAACACGACACATCTCTGACCTGGAACCTGAG 300
DB 2553 CGAGATGTTAGTGCACCCAGTACTGCAACACGACACATCTCTGACCTGGAACCTGAG 2612
QY 301 GTCCGAGAACATGATCATACCGAATACAACTGCTCAAGGTGCTGGACCTGGGCAATGC 360
DB 2613 GTCCGAGAACATGATCATACCGAATACAACTGCTCAAGGTGCTGGACCTGGGCAATGC 2672
QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTCAAGGACTACCTAGAGAC 420
DB 2673 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTCAAGGACTACCTAGAGAC 2732
QY 421 CATGGCTCCAGAGTCTCGAGGGCCAGGGGGCTTTCCACAGACAGACATCTGGGCGAT 480
DB 2733 CATGGCTCCAGAGTCTCGAGGGCCAGGGGGCTTTCCACAGACAGACATCTGGGCGAT 2792
QY 481 CGGTGTGACAGCTTTCATCATGCTGAGCGCCGAGTACCCGGTGAAGGAGGAGGTCGACG 540
DB 2793 CGGTGTGACAGCTTTCATCATGCTGAGCGCCGAGTACCCGGTGAAGGAGGAGGTCGACG 2852
QY 541 CGACTGACAGAGAGTCTCGAGGGGCTGGTCCGGCTGAGCCGCTGCTACCGGGGCT 600
DB 2853 CGACTGACAGAGAGTCTCGAGGGGCTGGTCCGGCTGAGCCGCTGCTACCGGGGCT 2912
QY 601 GTCCGGGGGGCGGCTGGCTTCTTCTGCGAGACATCTGTGCGCCCGAGCCCTGGGGCGGGCC 660
DB 2913 GTCCGGGGGGCGGCTGGCTTCTTCTGCGAGACATCTGTGCGCCCGAGCCCTGGGGCGGGCC 2972
QY 661 GTGGGCTCAGAGTCTGCTGAGTGCCTGCTGAACAGAGAGGGCGGCGCTGTTGCGG 720
DB 2973 GTGGGCTCAGAGTCTGCTGAGTGCCTGCTGAACAGAGAGGGCGGCGCTGTTGCGG 3032
QY 721 GCCCGCGCCGCTGACCTTCCCTACCGCGCGGCTCGGGTCTTCTGTCGCAATCGCGAGAA 780

DB 3033 GCCCGCGCGCTGACCTTCCCTACCGCGGGTGGCGGTCTTCTGTCGCAATCGCGAGAA 3092
QY 781 GAGACGGCGCTGCTGTACAGAGGACACAACTGGCCAGGTCGCTGAGGGTTCGCGCG 840
DB 3093 GAGACGGCGCTGCTGTACAGAGGACACAACTGGCCAGGTCGCTGAGGGTTCGCGCG 3152
QY 841 GCCACACCTTGGTCTTCCCGCTGGGGGTGCTGAGAGCGCCCAATAAAAAACGCCACGC 900
DB 3153 GCCACACCTTGGTCTTCCCGCTGGGGGTGCTGAGAGCGCCCAATAAAAAACGCCACGC 3212
QY 901 CGGCGAGAAAAA 913
DB 3213 CGGCGAGAAAAA 3225

RESULT 8
US-10-415-011-44
; Sequence 44, Application US/10415011
; Publication No. US20040053394A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: XU, Yuming
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: YAO, Monique G.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: DING, Li
; APPLICANT: TANG, Y. Tom
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: NGUYEN, Dannel B.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LU, Yan
; APPLICANT: YUE, Henry
; APPLICANT: BURFORD, Neil
; APPLICANT: BANDMAN, Olga
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: LAL, Preeti G.
; APPLICANT: RECIPON, Shirley A.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: SWARNAKER, Anita
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: KEAN, Farrán A.
; APPLICANT: ISON, Craig H.
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0262 USN
; CURRENT APPLICATION NUMBER: US/10/415,011
; PRIOR FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/47728
; PRIOR FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,410
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/244,068
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60/245,708
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/247,672
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/249,565
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,730
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 60/250,807
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PERL Program
; SEQ ID NO 44
; LENGTH: 5454

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20040053394A1 7638121CB1
US-10-415-011-44

Query Match 93.9%; Score 911.4; DB 13; Length 5454;
Best Local Similarity 99.9%; Pred. No. 2.3e-206;
Matches 912; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGCCAGATCATCCCTACCAACCCCAAGACAGACAGCAGT 60
DB 4542 CAGCGGGGGGGCTGGCGCCAGATCATCCCTACCAACCCCAAGACAGACAGCAGT 4601

QY 61 GCTGCGGAATACGAGGCGCTCAAGGGCTGGCGCCAGCCTGGCCAGCTGCAAGC 120
DB 4602 GCTGCGGAATACGAGGCGCTCAAGGGCTGGCGCCAGCCTGGCCAGCTGCAAGC 4661

QY 121 AGCTTACCTACGCGCCCGGACCTGGTCTCATCTTGGAGTGTGCTTGGCCCGAGCT 180
DB 4662 AGCTTACCTACGCGCCCGGACCTGGTCTCATCTTGGAGTGTGCTTGGCCCGAGCT 4721

QY 181 GCTCCCTCCCTGGCCGAGAGGGCTCCTACTCAGAAATCCAGGTGAAGACTACCTGTG 240
DB 4722 GCTCCCTCCCTGGCCGAGAGGGCTCCTACTCAGAAATCCAGGTGAAGACTACCTGTG 4781

QY 241 GCAGATGTTGATGCGCACCCAGTACTGCAAAACAGGACATCTGTCACCTGGACCTGAG 300
DB 4782 GCAGATGTTGATGCGCACCCAGTACTGCAAAACAGGACATCTGTCACCTGGACCTGAG 4841

QY 301 GTCCGAGAACATGATCATACCGAATAACAACCTGCTCAAGGTCTGGACCTGGGCAATGC 360
DB 4842 GTCCGAGAACATGATCATACCGAATAACAACCTGCTCAAGGTCTGGACCTGGGCAATGC 4901

QY 361 ACAGAGCTCTGCGCAGGAGAGGTGCTCCCTTCAGACAAGTTCAAGGACTACCTAGAGAC 420
DB 4902 ACAGAGCTCTGCGCAGGAGAGGTGCTCCCTTCAGACAAGTTCAAGGACTACCTAGAGAC 4961

QY 421 CATGCTCCAGAGCTCTGCGAGGCGCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480
DB 4962 CATGCTCCAGAGCTCTGCGAGGCGCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 5021

QY 481 CGGTGTGACAGCTTTCATCATGCTGAGCGCGAGTACCCGCTGAGCAGCGAGGTGCAAG 540
DB 5022 CGGTGTGACAGCTTTCATCATGCTGAGCGCGAGTACCCGCTGAGCAGCGAGGTGCAAG 5081

QY 541 CGACTGACAGAGAGTCTGCGAGGCGCTGGTCCGGCTGAGCGCGCTGCTACGCGGGCT 600
DB 5082 CGACTGACAGAGAGTCTGCGAGGCGCTGGTCCGGCTGAGCGCGCTGCTACGCGGGCT 5141

QY 601 GTCCGGGGCGCGGTGGCTTCTTCGCGAGCACTCTGTGCGCCAGCCCTGGGGCGGCG 660
DB 5142 GTCCGGGGCGCGGTGGCTTCTTCGCGAGCACTCTGTGCGCCAGCCCTGGGGCGGCG 5201

QY 661 CTGCGCGTCCAGCTGCTGAGTCCCGTAAACAGAGAGGCGCGCGCTGTTCCGG 720
DB 5202 CTGCGCGTCCAGCTGCTGAGTCCCGTAAACAGAGAGGCGCGCGCTGTTCCGG 5261

QY 721 GCCCGCGCGTGAACCTTCCCTACCGCGCGCTGGGCTTCTGTCGCGCATCCGAGAA 780
DB 5262 GCCCGCGCGTGAACCTTCCCTACCGCGCGCTGGGCTTCTGTCGCGCATCCGAGAA 5321

QY 781 GAGACGCGCTGCTGTAAAGAGGCAACAACCTGGCCAGGTGGCTGAGGGTCCGCCCG 840
DB 5322 GAGACGCGCTGCTGTAAAGAGGCAACAACCTGGCCAGGTGGCTGAGGGTCCGCCCG 5381

QY 841 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGAGACGCGCGCAATAAAGACCCAGC 900
DB 5382 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGAGACGCGCGCAATAAAGACCCAGC 941

QY 901 CGGCGCAGAAAA 913
|||||

Db 5442 CGGCGCAGAAAA 5454

RESULT 9
US-10-425-114-26278
; Sequence 26278, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 26278
; LENGTH: 1253
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4119-136-D3_FLI
US-10-425-114-26278

Query Match 93.3%; Score 906.4; DB 13; Length 1253;
Best Local Similarity 99.9%; Pred. No. 2.6e-205;
Matches 907; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGCCAGATCATCCCTACCAACCCCAAGACAGACAGCAGT 60
DB 346 CAGCGGGGGGGCTGGCGCCAGATCATCCCTACCAACCCCAAGACAGACAGCAGT 405

QY 61 GCTGCGCGAATACGAGGCGCTCAAGGGCTGGCGCCAGCCTGGCCAGCTGCGCACGC 120
DB 406 GCTGCGCGAATACGAGGCGCTCAAGGGCTGGCGCCAGCCTGGCCAGCTGCGCACGC 465

QY 121 AGCTTACCTACGCGCCCGGACCTGGTGTCTATCTTTGAGTGTGTCTTGGCCCGAGCT 180
DB 466 AGCTTACCTACGCGCCCGGACCTGGTGTCTATCTTTGAGTGTGTCTTGGCCCGAGCT 525

QY 181 GCTCCCTCCCTGGCGAGAGGGCTCTTACTCAGAAATCCAGGTGAAGACTACCTGTG 240
DB 526 GCTCCCTCCCTGGCGAGAGGGCTCTTACTCAGAAATCCAGGTGAAGACTACCTGTG 585

QY 241 GCAGATGTTGAGTGCACCCAGTACTGCAAAACAGACATCTTGCACCTGGACCTGAG 300
DB 586 GCAGATGTTGAGTGCACCCAGTACTGCAAAACAGACATCTTGCACCTGGACCTGAG 645

QY 301 GTCCGAGAACATGATCATCACCGAATACAACTGCTCAAGGTGCTGGACCTGGGCAATGC 360
DB 646 GTCCGAGAACATGATCATCACCGAATACAACTGCTCAAGGTGCTGGACCTGGGCAATGC 705

QY 361 ACAGAGCTCTCAGCCAGAGAGAGGTGCTGGCCCTCAGAAAGTTCAAGGACTACCTAGAGAC 420
DB 706 ACAGAGCTCTCAGCCAGAGAGAGGTGCTGGCCCTCAGAAAGTTCAAGGACTACCTAGAGAC 765

QY 421 CATGGCTCCAGAGCTCTCAGAGGCGCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480
DB 766 CATGGCTCCAGAGCTCTCAGAGGCGCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 825

QY 481 CGGTGTGACAGCTTTCATCATGCTGAGCGCGAGTACCCTGGTGGAGCAGGGTGCACG 885
DB 826 CGGTGTGACAGCTTTCATCATGCTGAGCGCGAGTACCCTGGTGGAGCAGGGTGCACG 885

QY 541 CGACTGACAGAGACTCTGCGAAGGGGCTGGTTCGGCTGAGCGGCTGTCTACGCGGGCT 600
DB 886 CGACTGACAGAGACTCTGCGAAGGGGCTGGTTCGGCTGAGCGGCTGTCTACGCGGGCT 945

QY 601 GTCCGGGGGGCGGTGGCTTCTTGGCAGACATCTGTGCGCGCCAGCCCTGGGGCGGCG 660

Db 946 GTCCGGGGCGCCGCTTCCCTGCGAGCACTCTGTGGCCCGAGCCCTGGGGCCGGCC 1005
Qy 661 CTGGCGCTCCAGCTGCGCTGAGTCCCGTGAAGTAAACAGAGAGGGCCCGCTGTTCCGG 720
Db 1006 CTGGCGCTCCAGCTGCGCTGAGTCCCGTGAAGTAAACAGAGAGGGCCCGCTGTTCCGG 1065
Qy 721 GCCCGCGCCGCTGACCTTCCCTACCGCGCGCTGCGCTCTTCGTGCGCAATCGCGAGAA 780
Db 1066 GCCCGCGCCGCTGACCTTCCCTACCGCGCGCTGCGCTCTTCGTGCGCAATCGCGAGAA 1125
Qy 781 GAGACGCGCTGCTGTAAAGAGGCAACCTGGCCCGAGTGGCTGAGGGTGGCCCG 840
Db 1126 GAGACGCGCTGCTGTAAAGAGGCAACCTGGCCCGAGTGGCTGAGGGTGGCCCG 1185
Qy 841 GCCACACCTTGTCTCCCGCTGGGGTGGCTGCGAGCGCCCAATAAAAAAGCCCGAGC 900
Db 1186 GCCACACCTTGTCTCCCGCTGGGGTGGCTGCGAGCGCCCAATAAAAAAGCCCGAGC 1245
Qy 901 CGGGCGAG 908
Db 1246 CGGGCGAG 1253

RESULT 10

US-10-182-243-14
; Sequence 14, Application US/10182243
; Publication No. US20040048310A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY D.
; APPLICANT: WHYTE, DAVID
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; APPLICANT: MARTINEZ, RICARDO
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES AND PROTEIN KINASE-LIKE
; TITLE OF INVENTION: ENZYMES
; FILE REFERENCE: 038602/1366
; CURRENT APPLICATION NUMBER: US/10/182,243
; CURRENT FILING DATE: 2003-07-07
; PRIOR APPLICATION NUMBER: PCT/US01/02337
; PRIOR FILING DATE: 2001-01-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 4936
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-182-243-14

Query Match 93.3%; Score 906.4; DB 13; Length 4936;
Best Local Similarity 99.9%; Pred. No. 3.5e-205;
Matches 907; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CAGCGGGCGGGCTGGCGGCAAGATCATCCCTACACCCCAAGCAAGACAGCAGT 60
Db 4029 CAGCGGGCGGGCTGGCGGCAAGATCATCCCTACACCCCAAGCAAGACAGCAGT 4088
Qy 61 GCTCGCGCAATACGAGGCGCTCAAGGGCTGCGCCACCCGACCTGGCCGAGTGCACGC 120
Db 4089 GCTCGCGCAATACGAGGCGCTCAAGGGCTGCGCCACCCGACCTGGCCGAGTGCACGC 4148
Qy 121 AGCTACTCTAGCCCCCGGACCTGGTCTCATTTGGAGCTGTGCTCTGGGCGCGAGCT 180
Db 4149 AGCTACTCTAGCCCCCGGACCTGGTCTCATTTGGAGCTGTGCTCTGGGCGCGAGCT 4208
Qy 181 GCTCCCTGCTGCGCGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGACTACCTGTG 240
Db 4209 GCTCCCTGCTGCGCGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGACTACCTGTG 4268
Qy 241 GCAGATGTTAGTGCCACCCAGTACCTGCAACACAGCACATCTCTGCACTTGAACCTTGA 300
Db 4269 GCAGATGTTAGTGCCACCCAGTACCTGCAACACAGCACATCTCTGCACTTGAACCTTGA 4328

Qy 301 GTCCGAGAACATGATCATCACCAATACAACTCTCAAGTCTGTGACCTGTGGCAATGC 360
Db 4329 GTCCGAGAACATGATCATCACCAATACAACTCTCAAGTCTGTGACCTGTGGCAATGC 4388
Qy 361 ACAGAGCTTCAGCAGGAGAGGTGCTGCTCTCAAGTCTCAAGTCTCAAGTCTCAAGTCT 420
Db 4389 ACAGAGCTTCAGCAGGAGAGGTGCTGCTCTCAAGTCTCAAGTCTCAAGTCTCAAGTCT 4448
Qy 421 CATGCTCTCAGAGCTTCCTGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480
Db 4449 CATGCTCTCAGAGCTTCCTGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 4508
Qy 481 CGGTGTGACAGCTTCATCATCTGAGGCGCGAGTACCCGCTGAGCAGGAGGTGACAG 540
Db 4509 CGGTGTGACAGCTTCATCATCTGAGGCGCGAGTACCCGCTGAGCAGGAGGTGACAG 4568
Qy 541 CGACCTGACAGAGAGTTCGCAAGGGGCTGCTCCGGCTGAGCGCTGCTACGCGGGGCT 600
Db 4569 CGACCTGACAGAGAGTTCGCAAGGGGCTGCTCCGGCTGAGCGCTGCTACGCGGGGCT 4628
Qy 601 GTCCGGGGCGCGCTGCGCTTCTGCGCAGCACTCTGTGCGCCGAGCCCTGGGGCCGGCC 660
Db 4629 GTCCGGGGCGCGCTGCGCTTCTGCGCAGCACTCTGTGCGCCGAGCCCTGGGGCCGGCC 4688
Qy 661 CTGCGCGTCCAGCTGCGCTGAGTCCCGTGAAGAGGAGGGCCCGGCTGTTCCGC 720
Db 4689 CTGCGCGTCCAGCTGCGCTGAGTCCCGTGAAGAGGAGGGCCCGGCTGTTCCGC 4748
Qy 721 GCCCGCGCGCTGACCTTCCCTACCGCGCGCTCGCGCTCTTCGTGCGCAATCGCGAGAA 780
Db 4749 GCCCGCGCGCTGACCTTCCCTACCGCGCGCTCGCGCTCTTCGTGCGCAATCGCGAGAA 4808
Qy 781 GAGAGCGCGCTGCTGTAAAGAGGCAACCTGCGCCAGTGGCTGAGGAGTGGCCCG 840
Db 4809 GAGAGCGCGCTGCTGTAAAGAGGCAACCTGCGCCAGTGGCTGAGGAGTGGCCCG 4868
Qy 841 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGCGAGCGCCCAATAAAAAAGCCCGAGC 900
Db 4869 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGCGAGCGCCCAATAAAAAAGCCCGAGC 4928
Qy 901 CGGGCGAG 908
Db 4929 CGGGCGAG 4936

RESULT 11

US-10-307-019-3
; Sequence 3, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenlin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SCIOS, INC.
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS-021DVI
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5007
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (97) ... (4926)
US-10-307-019-3

Query Match

93.2%; Score 905.4; DB 15; Length 5007;

| | | | | | | | | | |
|---|------|---|------|--|--|--|--|--|--|
| Best Local Similarity 99.9%; Pred. No. 6e-205; Matches 906; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | | | | | | | | |
| Qy | 1 | CAGCGGGCGGCGCTGCGCGCAGATATCCCTACACCCCAAGGACAGACAGT | 60 | | | | | | |
| Db | 4101 | CAGCGGGCGGCGCTGCGCGCAGATATCCCTACACCCCAAGGACAGACAGT | 4160 | | | | | | |
| Qy | 61 | GCTCGGGAATACAGAGGCTTCAAGGCGCTGCGCACCCGACCTGCGCCAGCTGCACGC | 120 | | | | | | |
| Db | 4161 | GCTCGGGAATACAGAGGCTTCAAGGCGCTGCGCACCCGACCTGCGCCAGCTGCACGC | 4220 | | | | | | |
| Qy | 121 | AGCTACTAGCCCGCGGACCTGCTGCTCATCTTGAGCTGCTGCGCCGAGCT | 180 | | | | | | |
| Db | 4221 | AGCTACTAGCCCGCGGACCTGCTGCTCATCTTGAGCTGCTGCGCCGAGCT | 4280 | | | | | | |
| Qy | 181 | GCTCCCTGCTGCGCAGAGGCGCTCTACTCAGAATCCAGAGTGAAGACTACCTGTG | 240 | | | | | | |
| Db | 4281 | GCTCCCTGCTGCGCAGAGGCGCTCTACTCAGAATCCAGAGTGAAGACTACCTGTG | 4340 | | | | | | |
| Qy | 241 | GCAGATGTGATGCGACCCAGTACTGTCACAAACAGACATCTGTCACCTGACCTGAG | 300 | | | | | | |
| Db | 4341 | GCAGATGTGATGCGACCCAGTACTGTCACAAACAGACATCTGTCACCTGACCTGAG | 4400 | | | | | | |
| Qy | 301 | GTCGAGAACATGATCATCCGGAATACAACTGCTCAAGTCTGAGACTGGGCAATGC | 360 | | | | | | |
| Db | 4401 | GTCGAGAACATGATCATCCGGAATACAACTGCTCAAGTCTGAGACTGGGCAATGC | 4460 | | | | | | |
| Qy | 361 | ACAGAGCCTCAGCCAGAGAGTGTGCTGCTCAGCAAGTTCAAGACTACCTAGAGAC | 420 | | | | | | |
| Db | 4461 | ACAGAGCCTCAGCCAGAGAGTGTGCTGCTCAGCAAGTTCAAGACTACCTAGAGAC | 4520 | | | | | | |
| Qy | 421 | CATGGCTCAGAGCTCTGAGGGCGAGGGGCTGTTCCACAGACAGACATCTGGGCGAT | 480 | | | | | | |
| Db | 4521 | CATGGCTCAGAGCTCTGAGGGCGAGGGGCTGTTCCACAGACAGACATCTGGGCGAT | 4580 | | | | | | |
| Qy | 481 | CGGTGTGACAGCTTCATCATGCTGAGCGCGAGTACCGGTGAGCAGCGAGGTGCACG | 540 | | | | | | |
| Db | 4581 | CGGTGTGACAGCTTCATCATGCTGAGCGCGAGTACCGGTGAGCAGCGAGGTGCACG | 4640 | | | | | | |
| Qy | 541 | CGACCTGACAGAGACTGCGCAAGGGGCTGGTCCGGCTGAGCCGCTGCTACGGCGGGCT | 600 | | | | | | |
| Db | 4641 | CGACCTGACAGAGACTGCGCAAGGGGCTGGTCCGGCTGAGCCGCTGCTACGGCGGGCT | 4700 | | | | | | |
| Qy | 601 | GTCGCGGGCGCGGTGCGCTTCTTCGCGAGCACTCTGTGCGCCCGAGCCCTGGGCGCGCC | 660 | | | | | | |
| Db | 4701 | GTCGCGGGCGCGGTGCGCTTCTTCGCGAGCACTCTGTGCGCCCGAGCCCTGGGCGCGCC | 4760 | | | | | | |
| Qy | 661 | CTGCGCGTCCAGCTGCTGCGAGTCCCGTGGCTAACAGAGAGGGCGCGCTGTTCGCG | 720 | | | | | | |
| Db | 4761 | CTGCGCGTCCAGCTGCTGCGAGTCCCGTGGCTAACAGAGAGGGCGCGCTGTTCGCG | 4820 | | | | | | |
| Qy | 721 | GCCGCGCCGCTGACCTTCCCTACCGCGGCTGCGCGTCTTCGTGCGCAATCGGAGAA | 780 | | | | | | |
| Db | 4821 | GCCGCGCCGCTGACCTTCCCTACCGCGGCTGCGCGTCTTCGTGCGCAATCGGAGAA | 4880 | | | | | | |
| Qy | 781 | GAGACCGCGCTGCTGTACAGAGGACAACTGCGCCCGAGGTGCGCTGAGGGTGGCCCGG | 840 | | | | | | |
| Db | 4881 | GAGACCGCGCTGCTGTACAGAGGACAACTGCGCCCGAGGTGCGCTGAGGGTGGCCCGG | 4940 | | | | | | |
| Qy | 841 | GCCACACCTTGGTCTCCCGCTGGGGTGGCTGCGAGCGCGCCAAATAAAGCCGCGCAGC | 900 | | | | | | |
| Db | 4941 | GCCACACCTTGGTCTCCCGCTGGGGTGGCTGCGAGCGCGCCAAATAAAGCCGCGCAGC | 5000 | | | | | | |
| Qy | 901 | CGGGCGA 907 | | | | | | | |
| Db | 5001 | CGGGCGA 5007 | | | | | | | |

RESULT 12
US-10-307-019-5
; Sequence 5, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:

| | | | | | | | | | |
|---|------|---|------|--|--|--|--|--|--|
| APPLICANT: Zeng, Wenlin APPLICANT: Stanton, Lawrence APPLICANT: SCIOS, INC. FILE REFERENCE: SCIOS 021DV1 CURRENT APPLICATION NUMBER: US/10/307,019 CURRENT FILING DATE: 2002-11-26 PRIOR APPLICATION NUMBER: 09/548,473 PRIOR FILING DATE: 2000-04-13 PRIOR APPLICATION NUMBER: 60/129,552 PRIOR FILING DATE: 1999-04-16 NUMBER OF SEQ ID NOS: 9 SOFTWARE: PastSeq for Windows Version 4.0 SEQ ID NO 5 LENGTH: 7928 TYPE: DNA ORGANISM: Homo sapiens FEATURE: NAME/KEY: CDS LOCATION: (60)....(7847) US-10-307-019-5 | | | | | | | | | |
| Query Match 93.2%; Score 905.4; DB 15; Length 7928; Best Local Similarity 99.9%; Pred. No. 6.e-205; Matches 906; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | | | | | | | | |
| Qy | 1 | CAGCGGGCGGCGCTGCGCGCAGATATCCCTACACCCCAAGGACAGACAGT | 60 | | | | | | |
| Db | 7022 | CAGCGGGCGGCGCTGCGCGCAGATATCCCTACACCCCAAGGACAGACAGT | 7081 | | | | | | |
| Qy | 61 | GCTGCGCGAATACAGAGGCGCTCAAGGGCTGCGCCACCCGACCTGCGCCAGCTGCACGC | 120 | | | | | | |
| Db | 7082 | GCTGCGCGAATACAGAGGCGCTCAAGGGCTGCGCCACCCGACCTGCGCCAGCTGCACGC | 7141 | | | | | | |
| Qy | 121 | AGCTACTAGCCCGCGGACCTGCTGCTCATCTTGAGCTGCTGCGCCGAGCT | 180 | | | | | | |
| Db | 7142 | AGCTACTAGCCCGCGGACCTGCTGCTCATCTTGAGCTGCTGCGCCGAGCT | 7201 | | | | | | |
| Qy | 181 | GCTCCCTGCTGCGCAGAGGCGCTTCTACTCAGAATCCAGAGTGAAGACTACCTGTG | 240 | | | | | | |
| Db | 7202 | GCTCCCTGCTGCGCAGAGGCGCTTCTACTCAGAATCCAGAGTGAAGACTACCTGTG | 7261 | | | | | | |
| Qy | 241 | GCAGATGTGATGTCGCCACCCAGTACTGTGCAACCAAGCACTCTGCACTGGACCTGAG | 300 | | | | | | |
| Db | 7262 | GCAGATGTGATGTCGCCACCCAGTACTGTGCAACCAAGCACTCTGCACTGGACCTGAG | 7321 | | | | | | |
| Qy | 301 | GTCGAGAACATGATCATCACCGAATACAACTGCTCAAGTCTCAAGTCTGGA | 360 | | | | | | |
| Db | 7322 | GTCGAGAACATGATCATCACCGAATACAACTGCTCAAGTCTGGA | 7381 | | | | | | |
| Qy | 361 | ACAGAGCCTCAGCCAGAGAGTGTGCTGCTCAGCAAAAGTTCAAGGACTACCTAGAGAC | 420 | | | | | | |
| Db | 7382 | ACAGAGCCTCAGCCAGAGAGTGTGCTGCTCAGCAAAAGTTCAAGGACTACCTAGAGAC | 7441 | | | | | | |
| Qy | 421 | CATGGCTCAGAGCTCTGAGGGCGAGGGGCTGTTCCACAGACAGACATCTGGGCGCAT | 480 | | | | | | |
| Db | 7442 | CATGGCTCAGAGCTCTGAGGGCGAGGGGCTGTTCCACAGACAGACATCTGGGCGCAT | 7501 | | | | | | |
| Qy | 481 | CGGTGTGACAGCTTCTCATCATGCTGAGCGCGGAGTACCCGGTGAAGGAGGAGTGCACG | 540 | | | | | | |
| Db | 7502 | CGGTGTGACAGCTTCTCATCATGCTGAGCGCGGAGTACCCGGTGAAGGAGGAGTGCACG | 7561 | | | | | | |
| Qy | 541 | CGACCTGCAAGAGGACTGCGCAAGGGGCTGCTGCGGCTGAGCGCTGTCTACCGGGGCT | 600 | | | | | | |
| Db | 7562 | CGACCTGCAAGAGGACTGCGCAAGGGGCTGCTGCGGCTGAGCGCTGTCTACCGGGGCT | 7621 | | | | | | |
| Qy | 601 | GTCGCGGGCGCGCTGCTGCTTCTGCGCAGCACTCTGTCGCCCGACCTTGGGGCGCGGC | 660 | | | | | | |
| Db | 7622 | GTCGCGGGCGCGCTGCTGCTTCTGCGCAGCACTCTGTCGCCCGACCTTGGGGCGCGGC | 7681 | | | | | | |
| Qy | 661 | CTGCGCGTCCAGCTGCTGCGAGTCCCGTGGCTTAACAGAGAGGGCGCGCTGTTCGCG | 720 | | | | | | |
| Db | 7682 | CTGCGCGTCCAGCTGCTGCGAGTCCCGTGGCTTAACAGAGAGGGCGCGCTGTTCGCG | 7741 | | | | | | |

721 GCGCGCGCCGCTGACCTTCCCTACCGCGCGCTGCGGCTCTTCTGGCGCAATCGCGAGAA 780
7742 GCGCGCGCCGCTGACCTTCCCTACCGCGCGCTGCGGCTCTTCTGGCGCAATCGCGAGAA 7801
781 GAGAGCGCGCTGCTGTACAGAGGCAACACTGCGCCAGTGCCTGAGGGTGCAGGCTG 840
7802 GAGAGCGCGCTGCTGTACAGAGGCAACACTGCGCCAGTGCCTGAGGGTGCAGGCTG 7861
841 GCCACACCTTGGTCTCCCGCTGCGGCTGCGTGCAGACGCGCCCAATAAATAAGCCCGCAGC 900
7862 GCCACACCTTGGTCTCCCGCTGCGGCTGCGTGCAGACGCGCCCAATAAATAAGCCCGCAGC 7921
901 CGGCGCA 907
7922 CGGCGCA 7928

RESULT 13
US-10-077-130-3
; Sequence 3, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCP1(N)
; CURRENT APPLICATION NUMBER: US/10/077,130
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-3

Query Match 85.4%; Score 829; DB 14; Length 7893;
Best Local Similarity 100.0%; Pred. No. 9e-187;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCGCGCGCGCTGCGCGCCAGATCATCCCTACCCAGGCAAGGCAAGACAGCAGT 60
DB 7065 CAGCGCGCGCGCTGCGCGCCAGATCATCCCTACCCAGGCAAGGCAAGACAGCAGT 7124

QY 61 GCTCGCGGAATACAGAGCGCTTAAAGGCGCTGCGCCAGCCAGCAGTGCAGCAGT 120
DB 7125 GCTCGCGGAATACAGAGCGCTTAAAGGCGCTGCGCCAGCCAGCAGTGCAGCAGT 7184

QY 121 AGCTTACTCAGCCCGCGGACCTGCTCATCTTGGAGCTGCTCTGGGCGCGAGCT 180
DB 7185 AGCTTACTCAGCCCGCGGACCTGCTCATCTTGGAGCTGCTCTGGGCGCGAGCT 7244

QY 181 GCTCCCTGCTGCGCGAGAGGCGCTCTACTCAGAAATCGAGGCTGAAGGACTACCTGTG 240
DB 7245 GCTCCCTGCTGCGCGAGAGGCGCTCTACTCAGAAATCGAGGCTGAAGGACTACCTGTG 7304

QY 241 GCAGATCTTGGTGCACCGGACCTGCTGCAACACGAGCAGCAGTGCAGCTGAGACCTGAG 300
DB 7305 GCAGATCTTGGTGCACCGGACCTGCTGCAACACGAGCAGCAGTGCAGCTGAGACCTGAG 7364

QY 301 GTCGAGAAATGATCATCAACCGAATCAACCTGCTCAAGGCTGCTGGAACCTGGGCAATGC 360
DB 7365 GTCGAGAAATGATCATCAACCGAATCAACCTGCTCAAGGCTGCTGGAACCTGGGCAATGC 7424

QY 361 ACAGAGCTCAGCAGAGAGAGTGTGCTGCTCAGCAAGATTCAGAGGACTACCTAGAGAC 420
DB 7425 ACAGAGCTCAGCAGAGAGAGTGTGCTGCTCAGCAAGATTCAGAGGACTACCTAGAGAC 7484

QY 421 CATGGCTCCAGAGCTCTCGAGGCGCGAGGCGGCTGTTCACAGACAGACATCTGGGCCAT 480

Db 7485 CATGGCTCCAGAGCTCTCGAGGCGCGAGGCGCTGTTCACAGACAGACATCTGGGCCAT 7544
QY 481 CCGTGTGACAGAGCTTATCATGCTGAGCGCCAGATACCGGTGAGCAGGAGGTGACAG 540
Db 7545 CCGTGTGACAGAGCTTATCATGCTGAGCGCCAGATACCGGTGAGCAGGAGGTGACAG 7604
QY 541 CGACCTGACAGAGAGTGCAGAGGCGTGTGCGGCTGAGCGCGCTGCTACGCGGGGCT 600
Db 7605 CGACCTGACAGAGAGTGCAGAGGCGTGTGCGGCTGAGCGCGCTGCTACGCGGGGCT 7664
QY 601 GTCGCGGCGCGCGCTGCGCTTCTGCGGAGCAGCTCTGTGCGGCGCAGCCCTGGGCGCGCC 660
Db 7665 GTCGCGGCGCGCGCTGCGCTTCTGCGGAGCAGCTCTGTGCGGCGCAGCCCTGGGCGCGCC 7724
QY 661 CTGCGCGCTCCAGCTGCGTGCAGTGCCTGCTAACAGAGAGGCGCGCGCTGTGTCGCG 720
Db 7725 CTGCGCGCTCCAGCTGCGTGCAGTGCCTGCTAACAGAGAGGCGCGCGCTGTGTCGCG 7784
QY 721 GCGCGCGCGCGTGCAGCTTCCCTACCGCGCGCTGCGGCTCTTCTGCGGCAATCGAGAA 780
Db 7785 GCGCGCGCGCGTGCAGCTTCCCTACCGCGCGCTGCGGCTCTTCTGCGGCAATCGAGAA 7844
QY 781 GAGAGCGCGCTGCTGTACAGAGGCAACACTGCGCCAGGTCGCTGA 829
Db 7845 GAGAGCGCGCTGCTGTACAGAGGCAACACTGCGCCAGGTCGCTGA 7893

RESULT 14

US-10-077-130-6
; Sequence 6, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCP1(N)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6

Query Match 85.4%; Score 829; DB 14; Length 23907;
Best Local Similarity 100.0%; Pred. No. 1.1e-186;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCGCGCGCGCTGCGCGCCAGATCATCCCTACCCAGGCAAGGCAAGACAGCAGT 60
Db 23079 CAGCGCGCGCGCTGCGCGCCAGATCATCCCTACCCAGGCAAGGCAAGACAGCAGT 23138

QY 61 GCTCGCGGAATACAGAGCGCTTAAAGGCGCTGCGCCAGCCAGCAGTGCAGCAGT 120
Db 23139 GCTCGCGGAATACAGAGCGCTTAAAGGCGCTGCGCCAGCCAGCAGTGCAGCAGT 23198

QY 121 AGCTTACTCAGCCCGCGGACCTGCTCATCTTGGAGCTGCTCTGGGCGCGAGCT 180
Db 23199 AGCTTACTCAGCCCGCGGACCTGCTCATCTTGGAGCTGCTCTGGGCGCGAGCT 23258

QY 181 GCTCCCTGCTGCGCGAGAGGCGCTCTACTCAGAAATCGAGGCTGAAGGACTACCTGTG 240
Db 23259 GCTCCCTGCTGCGCGAGAGGCGCTCTACTCAGAAATCGAGGCTGAAGGACTACCTGTG 23318

QY 241 GCAGATGTTGAGTGCACCGGAGTGCAGCAACAGCAGCAGTGCAGCTGACCTGGAGCTGAG 300
Db 23319 GCAGATGTTGAGTGCACCGGAGTGCAGCAACAGCAGCAGTGCAGCTGACCTGGAGCTGAG 23378

Db 121 CTCCTGGAGGGCGGCGCTGTTCCACAGACACATCTGGCCATCGGTGTGACAGCC 180
QY 494 TTTCATCATCTGAGCGCCGAGTACCCGGTGTGAGCGAGGGTGCA 538
Db 181 TTTCATCATCTGAGCGCCGAGTACCCGGTGTGAGCGAGGGTGCA 225

Search completed: September 20, 2004, 02:51:26
Job time : 505.587 secs

QY 301 GTCGAGACATGATCATACCGAATACAACTGCTCAAGGTCTGGACCTGGCAATGC 360
Db 23379 GTCGAGAACATGATCATACCGAATACAACTGCTCAAGGTCTGGCAATGC 23438
QY 361 ACAGAGCCTCAGCCAGAGAGAGTGTGCTCCCTCAGACAAATTCAGGACTACCTAGAGAC 420
Db 23439 ACAGAGCCTCAGCCAGAGAGAGTGTGCTCCCTCAGACAAATTCAGGACTACCTAGAGAC 23498
QY 421 CATGGCTCCAGAGCTCCTGAGGGCCAGGGGCTGTTCCACAGACACATCTGGGCCAT 480
Db 23499 CATGGCTCCAGAGCTCCTGAGGGCCAGGGGCTGTTCCACAGACACATCTGGGCCAT 23558
QY 481 CGGTGTGACAGCCTTCATCATGTGTGAGCGCCGAGTACCCGGTGTGAGCGAGGGTGCAAG 540
Db 23559 CGGTGTGACAGCCTTCATCATGTGTGAGCGCCGAGTACCCGGTGTGAGCGAGGGTGCAAG 23618
QY 541 CGACCTGTGAGAGAGACTGTGCGAGGGGCTGCTCCGCTGAGCGGCTGCTACCGGGGCT 600
Db 23619 CGACCTGTGAGAGAGACTGTGCGAGGGGCTGCTCCGCTGAGCGGCTGCTACCGGGGCT 23678
QY 601 GTCGCGGGGCGCGCTGGGCTTCTTCGCGAGCAGTCTGTGCGCCCGAGCCCTGGGGCCGGCC 660
Db 23679 GTCGCGGGGCGCGCTGGGCTTCTTCGCGAGCAGTCTGTGCGCCCGAGCCCTGGGGCCGGCC 23738
QY 661 CTGCGCGTCCAGTGTGCTGTGAGTGTGCTTAAACAGAGAGGGCCCGGCTGTTCGCG 720
Db 23739 CTGCGCGTCCAGTGTGCTGTGAGTGTGCTTAAACAGAGAGGGCCCGGCTGTTCGCG 23798
QY 721 GCGCGCGCGCGTACCTTCCCTACCGCGGCTGCGGCTCTTGTGCGCAATGCGGAGAA 780
Db 23799 GCGCGCGCGCGTACCTTCCCTACCGCGGCTGCGGCTCTTGTGCGCAATGCGGAGAA 23858
QY 781 GAGAGCGCGCTGTGTGTACAGAGGCAACACTGGCCAGGTGCGCTGA 829
Db 23859 GAGAGCGCGCTGTGTGTACAGAGGCAACACTGGCCAGGTGCGCTGA 23907

RESULT 15

US-09-864-408A-2403
; Sequence 2403, Application US/09864408A
; Publication No. US20040009474A1el Human Polynucleotides and Polypeptides Enc
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1el Human Polynucleotides and Polypeptides Enc
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2403.
; LENGTH: 225
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-408A-2403

Query Match 23.0%; Score 223.4; DB 11; Length 225;
Best Local Similarity 99.6%; Pred. No. 2.1e-43;
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 314 ATCATCACCAGATACAACTGCTCAAGGTCTGGACCTGGGCAATGACAGAGCCTCAGC 373
Db 1 ATCATCACCAGATACAACTGCTCAAGGTCTGGACCTGGGCAATGACAGAGCCTCAGC 60
QY 374 CAGGAGAAGGTGTGCTCCCTCAGACAAATTCAGGACTACCTAGAGACCATTGCTCCAGAG 433
Db 61 CAGGAGAAGGTGTGCTCCCTCAGACAAATTCAGGACTACCTAGAGACCATTGCTCCAGAG 120
QY 434 CTCTGTGAGGGCCAGGGGCTGTTCCACAGACAGATCTGGGCCATCGGTGTGACAGCC 493

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 14:16:08 ; Search time 1977.82 Seconds
(without alignments)
10628.572 Million cell updates/sec

Title: US-10-077-130-4_COPY_3662_4146

Perfect score: 485

Sequence: 1 gacgtgtacaagacggga.....ggagcagctggcacacagga 485

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167515995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hcg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|-----------|--------------------|
| 1 | 485 | 100.0 | 20435 | 9 | HSJ2535 | AJ002535 Homo sapi |
| 2 | 279.4 | 57.6 | 19475 | 9 | HSJ314900 | AJ314900 Homo sapi |
| 3 | 279.4 | 57.6 | 135964 | 9 | AL353593 | AL353593 Homo sapi |
| 4 | 220.8 | 45.5 | 24545 | 9 | HSJ314908 | AJ314908 Homo sapi |
| 5 | 218.6 | 45.1 | 2488 | 6 | AX883644 | AX883644 Sequence |
| 6 | 218.6 | 45.1 | 2488 | 6 | BD160445 | BD160445 Primer fo |
| 7 | 218.6 | 45.1 | 2488 | 9 | AK024186 | AK024186 Homo sapi |
| 8 | 218.6 | 45.1 | 174612 | 2 | AC023889 | AC023889 Homo sapi |
| 9 | 216 | 44.5 | 10091 | 9 | HSJ314901 | AJ314901 Homo sapi |
| 10 | 211.2 | 43.5 | 3956 | 9 | AK128447 | AK128447 Homo sapi |
| 11 | 209.6 | 43.2 | 6400 | 9 | HSJ314903 | AJ314903 Homo sapi |
| 12 | 207.6 | 42.8 | 211829 | 10 | AL645854 | AL645854 Mouse DNA |
| 13 | 207.4 | 42.8 | 260998 | 2 | AC099089 | AC099089 Rattus no |
| 14 | 201.2 | 41.5 | 164766 | 2 | AC026657 | AC026657 Homo sapi |
| 15 | 201 | 41.4 | 44358 | 2 | AC142478 | AC142478 Rattus no |
| 16 | 201 | 41.4 | 244553 | 2 | AC098133 | AC098133 Rattus no |
| 17 | 180.2 | 37.2 | 25087 | 2 | AC096931 | AC096931 Rattus no |
| 18 | 136.8 | 28.2 | 526 | 11 | BV077840 | BV077840 S208F6660 |
| 19 | 74.4 | 15.3 | 5382 | 9 | AB014557 | AB014557 Homo sapi |
| 20 | 73.4 | 15.1 | 933 | 9 | BC061909 | BC061909 Homo sapi |
| 21 | 73.4 | 15.1 | 1704 | 9 | AF035292 | AF035292 Homo sapi |
| 22 | 73.4 | 15.1 | 3316 | 9 | BC007201 | BC007201 Homo sapi |
| 23 | 71.8 | 14.8 | 2025 | 9 | AK025946 | AK025946 Homo sapi |
| 24 | 70.4 | 14.5 | 5314 | 10 | AX122340 | AX122340 Mus muscu |
| 25 | 67.8 | 14.0 | 724 | 6 | AX870351 | AX870351 Sequence |
| 26 | 67.8 | 14.0 | 724 | 6 | BD150413 | BD150413 Primer fo |
| 27 | 67.8 | 14.0 | 2184 | 6 | AX883309 | AX883309 Sequence |
| 28 | 67.8 | 14.0 | 2184 | 6 | BD160248 | BD160248 Primer fo |
| 29 | 67.8 | 14.0 | 2184 | 9 | AK023854 | AK023854 Homo sapi |
| 30 | 67.6 | 13.9 | 3015 | 10 | BC056205 | BC056205 Mus muscu |
| 31 | 66 | 13.6 | 2417 | 10 | BC051643 | BC051643 Mus muscu |
| 32 | 65.8 | 13.6 | 2534 | 6 | AX834822 | AX834822 Sequence |
| 33 | 65.8 | 13.6 | 2534 | 9 | AK097489 | AK097489 Homo sapi |
| 34 | 65.8 | 13.6 | 2761 | 9 | BC043916 | BC043916 Homo sapi |
| 35 | 63.6 | 13.1 | 5390 | 9 | HSJ314904 | AJ314904 Homo sapi |
| 36 | 63.4 | 13.1 | 7218 | 6 | I66494 | I66494 Sequence 14 |
| 37 | 59.8 | 12.3 | 34667 | 9 | HSJ314905 | AJ314905 Homo sapi |
| 38 | 59.8 | 12.3 | 164766 | 2 | AC026657 | AC026657 Homo sapi |
| 39 | 59.8 | 12.3 | 174612 | 2 | AC023889 | AC023889 Homo sapi |
| 40 | 57.4 | 11.8 | 572 | 6 | AX576249 | AX576249 Sequence |
| 41 | 56.4 | 11.6 | 95745 | 9 | AL359510 | AL359510 Human DNA |
| 42 | 55.8 | 11.5 | 225471 | 2 | AC112361 | AC112361 Rattus no |
| 43 | 55.8 | 11.5 | 266167 | 2 | AC121633 | AC121633 Rattus no |
| 44 | 54.8 | 11.3 | 187999 | 2 | AC115011 | AC115011 Mus muscu |
| 45 | 54.6 | 11.3 | 135033 | 9 | AC009955 | AC009955 Homo sapi |

ALIGNMENTS

RESULT 1
HSAJ2535
LOCUS
DEFINITION Homo sapiens mRNA for obscurin (OBSN gene).
ACCESSION AJ002535
VERSION AJ002535.1 GI:15026973
KEYWORDS
SOURCE Homo sapiens (human).
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young, P., Ehler, E. and Gautel, M.
TITLE Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly
PRI 14-SEP-2001

JOURNAL J. Cell Biol. 154 (1), 123-136 (2001)
 MEDLINE 21342081
 PUBMED 11448995
 REFERENCE 2 (bases 1 to 20435)
 AUTHORS Gautel, M.S.
 TITLE Direct Submission
 JOURNAL Submitted (31-OCT-1997) Gautel M.S., Structural Biology Division,
 European Molecular Biology Laboratory, Meierhofstr. 1, Heidelberg,
 69117, GERMANY
 REMARK Revised by author 20-JUL-2001
 FEATURES Location/Qualifiers
 source
 1..20435
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1q42.1"
 /cell_type="cardiomyocyte"
 /tissue_type="cardiac"
 /dev_stage="adult"
 1..20435
 /gene="OBSCN"
 72..19934
 /gene="OBSCN"
 /function="titin binding"
 /codon_start=1
 /evidence=experimental
 /product="obscurin"
 /protein_id="CAC44768.1"
 /db_xref="GI:15026974"
 /db_xref="GOA:O96AA2"
 /db_xref="SPTREMBL:O96AA2"
 /translation="MDQPPSGAPRLTRPKAFVSVGKDATLSCOIIGNPTPOVSE
 KDQPVTAGFRLLAQGDLYRLTLDLQDGSQVCRANWAGFAAFAVGLVDVAE
 AACCAQFELLRLPTSLRVREGEATFCRVGSPAPVSNWAGRLGAEFGDPRVR
 BELGEALRLRAARPDGGTYEVRANPLGASAAAALVYDSDADATSRGTSTAA
 LLAHLQRREAMEAGAPASPPSTGTCTCTCTGEGKARLSYCVYTGPKETVWKGQ
 LVTEGRVYVDAQENFLVKILPKOSDGLTCTASNLVGYTSGVLVWREPAVP
 PKRLDLVREKESATFLCEVOPSTEAAMFKEETRLWASAYGTEEGTETRLVR
 NVSDADAVYICETPEGSRTVAELAVQGNLLRLPKRTAVRGDTAMFCVELAVPVP
 VHLRQEEVVAGRVSAISAEGRTHLITISQCCLEDVGVQVAFWAGDCQISTRCVSA
 KRLPLQPDVVPVARMSSVILSWSPPPHGERPVTIDGLVLEKKKLGIYTRICHEA
 EWAPTELTVADEENGFQFVSALNSFGQSPYLEFPFGTVHLAPKXALVRLPLKVA
 VEGEVTFVSDLTVASAGFWLDQALNASVYEIHCDDTRHTLTIREVPLSHGAQL
 KFVANGTESIRMEVRAAPGLTANKPPAAAREVRLARLHEEAOLLAELSDQAAATWL
 KDGTLSGPKYEVQASAGRRVLLVRDVARDDAGLYECVSRGGRIAVOLSVQGLARFL
 HKMAGSCDVAAGGPAQFCETSEAHVHVHVHKDGMELHSGERFLQEDVGTRELV
 AATVTRDDEGTSCRVGEDSVDFRLRVSEKVPFAKEQLARRKLAQEAASITLSEV
 AQTQETVTKDGLVSSSKVCEATGCTRRLLVQOQADAGEYSCBAGGQRLSFH
 LDVKEPFWFAQVQAHSEYQAEAGANATLSCVEAQAQAEVMMYKDGKLSLKVYH
 EAKCRRLLVQOQAGTDAGYCEARGQVSRFLHITPFPOMFAKEQSVHNEVQAE
 GASAMLSCVEAQAQETVTKDGLVSSSKVGMVKGCTRRLLVQOQADAGEYSC
 BAGQVSYFHLITTEPKGVFAKEQSVHNEVQAEAGTTAMLSCVEAQAQETVTKYKDGK
 KLSSSKVRMEVKGCTRLVQOQKADAGEYSCBAGGQVSRFLHITTEPKAVFAKEQ
 LVHNEVTEAGASATLSCVEAQAQETVTKDGLVSSSKVRIEAGCVRQLVQQA
 QVADAGEYTCAGQRLSHFLVSEKAVFAKEQLARRKVAQAEAIATLSCVEAQAQ
 TEVTKYKDGKLSLSSKVRMEVAGCTRRLLVQOQADAGEYSCBAGGQRLSFHLDVA
 EPKVVFAKEQFVHNEVQAGATLSCVEAQAQETVTKYKDGKLSLSSKVRMEVAG
 CTRLLVQOQADAGEYSCBAGGQRLSHFLHVAEPKAVFAKEQSPASREVQAEAGTSA
 TLSCVEAQAQETVTKYKDGKLSLSSKVRMEVAGCTRRLLVQOQADAGEYSCBAG
 QRLSHFLHVAEPKAVFAKEQPAHREVQAEAGATLSCVEAQAQETVTKYKDGKLSL
 SSKVRMEVAGCTRRLLVQOQADAGEYSCBAGGQRLSHFLHVAEPKAVFAKEQ
 PLVVKHEDLITATLTPSAATVTKOGVEIRRRSKRRLHVAELPEQISERPCHRE
 DSATYCRVGAEGDFPVVEEVAKFCLLEPVCVGGELGVTILACELSPAFAEVMVR
 CGNTQPVGRFQVMAEGPVRLTILGLRAEDAGEYSCSRDDHTSAQTSVPRVYK
 FMSGLTVAEEGEATFQCVSPSVAVVWFDGALLQSEKFAISQSGASHSLTIS
 DLVLEDAQOITVAEAGSSAALRVREAPVLPFKKLEPOTVEERSVTLVLESLTRWP
 ELKRTNRATALAPKQVIEHAEGARHLVLHNVGFADRGFFCGCTPDDTKTAKLTVM
 ROVLVRGLQVAREOGCTATMEVQLSHADVDSWTRDGLRFOQGTCTHLVAGPMHT
 LITSLGPEDSGVFAEGVTHSARLVITELPVPSPQDQVTLVTEKEKVLCELS
 RPNVDWMLKDXGLRAGKATMAAAGACRSLTIRYCEFADQGVVYCDADHAGSSAV
 KVQRTYTLTYRVLRAEDAGEIQAFAENAESRQLRVKELPVLVRLPKDAMEKXHR
 GVLEQCVSRASAAQVWFVFKSQELQPKQVYELVSDGLYRKLIIISDVHAEDEDTYTCDA
 DVKTSAQFFVEEQSIIVRGLQDVTVMPEAPAFECETSIIPSVRPPKMLLCKTVOAQ
 GNVGLEQGTVHRLMLRRTCTMTGTPVHTVKGSRSLRVSDIPVLTPLPPLKTCG
 RELQSVLSCDFRPAPKAVQKDDTPLSPSEKFMKSEGCMAELRLRLMPADAGVY
 RCOAGSASSTEVTVAREAVTGTPLQDABATEEGWASFCELSHEDEBEWLSNGMP
 LYNDSFHEISHKGRHTLVLSIQRADAGIVRASLKVSTARSAREVVKPVFLKALD
 DLASERGTLLALQCEVPEAHVVRKQDQVQGFSDKYDFHHTAGTGLVHVDVSPED
 AGLYTCYHGSBETARVVRVHLHVTIKRLKTMVELEGESCFECVLISHESASPAW
 TVGKTKWSSFRQATQGRKRIILVREAPASDAGEVVFVSRGLTYSKASITVRRPAA
 IIKPLEDDQVAGFDELRCBLSRAGTPVHMLKDKRKAKIRKSGTDDVYCEGTMLMVR
 GASLKADAGEYTCVEASKSTASLHKEKANCITEELTNLOVEEKGTAVFTCKTHPAA
 TVTWKGLLELRASCKHOPSOEGLTKLITSALEKADSDTVCIDIGAOQASRAOLVVG
 RVVHILEDLDVYQEGSSATFCRI SPANVEPVHMLDKTPLHANSLEIDAQPGGY
 HVLTLRQALKDSGTIIPFAGDQASAAALVTEKXSVFSRELTDATITEGEDTLVGE
 TSTCDIPMCVTKDGTURGSARCOLSHGHRAQLITGATLQDSGRYKCBAGGACSS
 IYVHAREVRFQELKOLEVLEGGATLRCVLSVAAPVKWCYGNVLRPDGKYSLRQ
 EGAMLELVNRLFPQDSGRYSCFQDQTTSATLTVTALPAQFIGKLRNKEATEGAT
 LRCLSTAPVWRKRGSETLRDGRYCLRDQGMCELCQIRGLAMVDAEYSCVCGEER
 TSASLTIRPMAHFIGRLHODESIEGATATLRCELSAAAPVWEKGRSLESDGRHSL
 RODGAVCLOICGLAVADAGEYSCVCGEERTSATLVKALPAKTEGLRNEEA VEGAT
 AMLWCELSKVPVWRKGPENLRDGRYILRQEGTRCELQICGLMADAGEYLCVCCQ
 ERTSATLITRALPARFIEDVQREAGATVLOCELSNAPVWRKSGSETLRDGRY
 SLRQDGTCELQIRGLAMADTGYSCVCGEERTSMTLTVRALPIKFTGLRNESEATEG
 ATAVRLCELSKVPVWRKGMHETLRDGRHSLRQDGCARCELQIRGLVADAGEYLCMC
 GKERTSMTLVRAPSKFIEGLRNEATEGDTATLWCELSKAAAPVWRKGMHETLRDGR
 RHLRQDGRSCELQIRGLVADAGEYSCVCGEERTSATLVKALPAKTEGLRNEEA VEGAT
 EGATVLOCELSKAAAPVWRKGMHETLRDGRHSLRQDGCARCELQIRGLVADAGEYLC
 VCGERTSATLITVRAPQVFRPELQSLQAESEGTATLQCELSPTATVWWSKGLQQLQ
 ANGRREPLQTAELVQLQDREDTGYCTCGSQATSATLTVTAAPVRLRELQHQ
 DYDEGTAHLCELSRAGASVWRKGLSLFPCKAKYQVQDGAALAEILLVGEVEDAG
 EVDCTGHTQSVASLSVVRPFRKTLQSLQETGDIARLCCOLDAESGAVQVWLK
 EGVHLHAGPKYEMRSQAGATRELLIHLQEAQDTGEYAVCTGQKTAASLVRTEPVTV
 RGLVDAETADEVEFCEVSRAGATVQVCLQGLPQOSNEVTVAVDRGRIHTLRUK
 CVTPEDAGTVFHLGNHSSAQLTVRAPENVTLLEPLOVOLSEGODASFOCLRSSAG
 QEARWALGVPLQANEMNDITVEQGTLLHLTLHKVTLLEDAGTVSFHVTSSSAQLKV
 TAKNTVGLNELEALGEALFECLQSQEVAHAHTWLLDDEPVTSENAEVFFENG
 LRHLLNLKPLQDPSQVTRFLAGMDVTSAFTVAGWRLEILEPLKNAARVAGQAQRT
 CTLSEAPVGEASWYINGAAVQDDSDVTADGSHQALLRSAPHQHAGEVTFACRD
 AVASARLGLGLPDDAEVAHSHSTVLSMAAPSDGGGLCGVRYRVEKAGATPQ
 WRLCHELVPGCEVVDGLAPEGTYRFRVAAVPGVAGEPVHLPGTVLAEPKVPVPPQ
 PASEQVAGAGEVLSLELVAAEAGVIVHKGMERIQGRGFVWWSQGMQLVIGK
 FTADQGEYHGLQAGSICPAAATFQVALSPASVDEAPQSLPPEAAQEGDLHLWEA
 LARKRREPTLDSISLSELPEDRSQRLQAEVAPDLSSEGTSTADELALGDADL
 SHTSSDDESRAGTSLVTLKAGRPGTSLASKVGAAPSPVKPQQQPLAAARPL
 LGDSTKDLGDSMDKAAVKITQAAFKYKVRKEMKQEGPMFHTGDTGFAQVGDALR
 LECVASKADVRAASLKDGBELTDGRHHHIDQDGTCSLLIAGLRADAGCYTCQVS
 NKFGVTHSACVVVSGSESAESSGELDDAPRAARLRLHRLFTSKSPAEVDEBLF
 VEQGRVEMCISKETAPVVPPEPLSLTSDAAPVFLTELQNEQVQDGPVPSFCV
 VTGPMPEVVRKQGLLEDDHYNEDQGGHLLITAVPADMGVYRCLANESNG
 BLSTKAEPLVLTSTYTDADATSESSYFSAQYLSRSEQEGTETDGGQLPVVVE
 ESRDLQVAPLTKLQFKVGPAPRLYMKDQQLTASAHIMTDLKLTLEIIS
 VTRDSDGYAAIISNAGAAVSARLLVRGDEPEEKPSADVHEQLVPPHMLERFTPK
 KVKSGSTTFEVKVEGVPVTVHMLEEAGEVLIWIGDTPGTVASSAACHLSLLLD
 VGRQHTYTCIASNAGQALCSASLHVSGLPKVEOEKVEKALISTFLOGTQAIISA
 QGQTASPADLGGQRKEEPLAALHLSLAEEVTEEFKLTSITQITMVSATLQA
 KLVPGGSDSDSKTPSPASPHGRSPSSIQSESSSEDDGDARGEIDFYVVTADLV
 PLGAQDAITLREGQVVEVLDAAHPLRLVTRKTSKSPRQGVSPAYLDLRLSP
 EWGAAPAEFPEGEASEDEYKARLSVILQELSEQAFVEELOQLSHQLHLRCPH
 VPIAVAGKAVIIFRNVDIGRHSFQLQELQCDTDDVAMCTIKNOAAFEQVLEFV
 GRVASEVYSTAQEFYKABEALLAGDPSPPPLOHLYLEQVQRYOYQALIK
 ELINARKNRCNCALLBQAVAVSALPQRAENLHVSIMENYPCITLQALGEPTRQGH
 IHWEGAGARNPMPKGNHRLVFNHLYICKPRDRSDTDTVSVYFRNMKLSIDLND
 QVEGDDARVWQREDRSVKLILQARTAIKSWKEICIQRLRLKAPVWRPDPFE
 ELADCTBELGTVKLACRVGTGPKVPSWKDGVKAVQVDPHLLIEDPDGSCALIDS
 LTVGDSQMGTFAAASAGNCTGLKILVQVPFVKVLSAFAVVEGDAQDFTCTIEGA
 PYPQIYVQKCALLITGNKFTLSEPSRGLLVLVRAASKEDLGLCELVNLRGSLAR
 ASALRLTQSMPLQAOCHREQLVAVEVTEQTKPKKTVITEETTVVKSPPRGOR
 RPSKSPSRSPSCASPLRGLLAPLLYLRGAGQPRRPEAPGPKVPTLYVTEA
 BAHSPALPLGSPQPKVVEETIEVWKMGQGVSPTEVPRSSGHLFTLPGATP
 GCDPNSNNNNKLLQAQAGGTWAGVREPLVFRVDARSVDAWASCMGSLBEGTMM
 EAGEEGEDGDFAVTEESQDTHSLGRDPRKILTHNGRMUTLADLEDYVGEGETPHC
 BGGQGFADPPCEVSVIQRIEGTPVDSVHGLGPRGLFRPGEAGSPF
 GPCVRSLEGTSFLLREAPARPVGSAPWTSQFCTRIIRSADSGOSFTTSLTSTQVTFG

| | | | |
|-----------------------|---|--------|---|
| ORIGIN | TVGIVTVLHICPDRDRGDEAAQP" | mrna | join(<4472..4747,5585..5860,6702..6977,7817..8092,11254..11529,17940..>18215) |
| Query Match | 100.0%; Score 485; DB 9; Length 20435; | | /gene="OBSCN" |
| Best Local Similarity | 100.0%; Pred. No.1e-99; | CDS | join(<4472..4747,5585..5860,6702..6977,7817..8092,11254..11529,17940..>18215) |
| Matches 485; | Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | /gene="OBSCN" |
| | | | /codon_start=3 |
| QY | 1 GACGTGGTACAGAGCGGAGAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAA | | /product="obscurin" |
| Db | 3662 GACGTGGTACAGAGCGGAGAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAA | | /protein_id="CAC85747.1" |
| | | | /db_xref="GI:21104330" |
| QY | 61 GGGCTGCACAGAAAGCGCTGGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGGAGTACAG | | /db_xref="SPTREMBL:Q8NH7" |
| Db | 3722 GGGCTGCACAGAAAGCGCTGGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGGAGTACAG | | /translation="PKVVPKQDVAAHSEVOAEAGASATLSCVEAQAQTEVMVYKDGKK |
| QY | 121 CTGCAGAGCTGGGGCCAGAGAGTCTCTTTCAACTGCAATCAAGAGCCCAAGGCGAGT | | LSSLVKHVVEAKGCRRLVVOAQKTDAGDYSCEARGQVRSRLHITEPKVYFAKEQS |
| Db | 3782 CTGCAGAGCTGGGGCCAGAGAGTCTCTTTCAACTGCAATCAAGAGCCCAAGGCGAGT | | VNEVOABAGAGAMLSCEVAQAQTEVTHYKDKKKLSSSKVGMVKGCTRRLLVLPQAG |
| QY | 181 GTTTCGCCAAGAGCAGTGTGGTGCAATAGTGGTGGGACTCAGCAGGGGCGAGTGGCCAC | | KADAGYCEAGGQQRVSFHLHITEFKGVFAKQSVRNEVOAEGATTAMLSCEVAQAQTE |
| Db | 3842 GTTTCGCCAAGAGCAGTGTGGTGCAATAGTGGTGGGACTCAGCAGGGGCGAGTGGCCAC | | EVTVYKDKKKLSSSKVRNEVKGCTRRLLVVOQVKADAGEYSCGAGGQQRVSFQHLITE |
| QY | 241 ACTGAGCTGTGAGTGGCCCGAGCCAGACAGAGTGACGTGGTACAGGATGGGAAGAA | | PKVFAKEQLVHNEVTRTEAGASATLSCVEAQAQTEVTHYKDKKKLSSSKVRIEAGC |
| Db | 3902 ACTGAGCTGTGAGTGGCCCGAGCCAGACAGAGTGACGTGGTACAGGATGGGAAGAA | | MRQLVVQQAQDAGEYTCGAGQRLSFHLDSVEPKAVFAKEQLAHRKQVQABAGATAT |
| QY | 301 GCTGAGCTCCAGTTCGAAGTGGCATAGAGGCTGCGGGCTGCATGCGGCGAGCTGGTGGT | | LSCVEAQAQTEVTHYKDKKKLSSSKVRMEAVGCTRRLLVVOACQADTGEYSCEAGGQ |
| Db | 3962 GCTGAGCTCCAGTTCGAAGTGGCATAGAGGCTGCGGGCTGCATGCGGCGAGCTGGTGGT | | RLSFSLDVAEPKVVPFAKEQPVHREVOAQAQSTTLSCVEAQAQTEVMVYKDKKKLSFS |
| QY | 361 GCAGCAGGCGAGCCAGGCGAGATGCTGGGAGGTACACCTGTGAGGCTGGGGCCAGCGGCT | | SKVRMEAVGCTRRLLVVOQAQVAGEYSCGAGSQRLSFHLVA" |
| Db | 4022 GCAGCAGGCGAGCCAGGCGAGATGCTGGGAGGTACACCTGTGAGGCTGGGGCCAGCGGCT | exon | <4472..4747 |
| QY | 421 CTCCTTCACCTGAGTGTTCAGAGCCCAAGCGGTGTTTCCAAAGAGAGCTGGGACACA | | /gene="OBSCN" |
| Db | 4082 CTCCTTCACCTGAGTGTTCAGAGCCCAAGCGGTGTTTCCAAAGAGAGCTGGGACACA | intron | /notes="all" |
| QY | 481 CAGGA 485 | | 4748..5584 |
| Db | 4142 CAGGA 4146 | | /gene="OBSCN" |
| | | exon | /note="a11" |
| | | | 5585..5860 |
| | | | /gene="OBSCN" |
| | | intron | /note="a12" |
| | | | 5861..6701 |
| | | | /gene="OBSCN" |
| | | exon | /note="a12" |
| | | | 6702..6977 |
| | | | /gene="OBSCN" |
| | | intron | /note="a13" |
| | | | 6978..7816 |
| | | | /gene="OBSCN" |
| | | exon | /note="a13" |
| | | | 7817..8092 |
| | | | /gene="OBSCN" |
| | | intron | /note="a14" |
| | | | 8093..11253 |
| | | | /gene="OBSCN" |
| | | exon | /note="a14" |
| | | | 11254..11529 |
| | | | /gene="OBSCN" |
| | | intron | /note="a15" |
| | | | 11530..17939 |
| | | | /gene="OBSCN" |
| | | exon | /note="a15" |
| | | | 17940..>18215 |
| | | | /gene="OBSCN" |
| | | | /note="a16" |
| ORIGIN | | | |
| Query Match | 57.6%; Score 279.4; DB 9; Length 19475; | | |
| Best Local Similarity | 99.6%; Pred. No.3.5e-53; | | |
| Matches 280; | Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | |
| QY | 163 CACAGAGCCCAAGGCGAGTGTTCGCAAGGAGCAGTTGGTCATATGAGGTGGGACTGA | | 222 |
| Db | 7812 CTCAGAGCCCAAGGCGAGTGTTCGCAAGGAGCAGTTGGTCATATGAGGTGGGACTGA | | 7871 |
| QY | 223 GGCAGGGGCGAGTGGCCACACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGCTG | | 282 |
| Db | 7872 GGCAGGGGCGAGTGGCCACACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGCTG | | 7931 |
| QY | 283 GTACAAGGATGGAGAAGCTGAGCTCCAGTTTCGAAGTGCATAGAGGCTGCGGGCTG | | 342 |
| Db | 7932 GTACAAGGATGGAGAAGCTGAGCTCCAGTTTCGAAGTGCATAGAGGCTGCGGGCTG | | 7991 |
| QY | 343 CATGCGGCGAGTGGTGGTGCAGCAGGCGAGGCCAGATGCTGGGGAGTACACCTGTGTA | | 402 |

| | |
|--------|--|
| intron | /gene="OBSCN" /number=5 593. .947 /gene="OBSCN" /number=5 948. .1256 /gene="OBSCN" /number=6 1257. .1719 /gene="OBSCN" /number=6 1720. .1986 /gene="OBSCN" /number=7 1987. .2577 /gene="OBSCN" /number=7 2578. .2859 /gene="OBSCN" /number=8 2860. .3169 /gene="OBSCN" /number=8 3170. .3445 /gene="OBSCN" /number=9 3446. 5480 /gene="OBSCN" /number=9 5481. .5750 /gene="OBSCN" /number=10 5751. .10615 /gene="OBSCN" /number=10 10616. .10891 /gene="OBSCN" /number=11 |
| exon | |
| intron | |
| exon | |
| intron | |
| exon | |
| intron | |
| exon | |
| intron | |
| exon | |
| intron | |
| exon | |

| Query Watch | 45.5%; | Score 220.8; | DB 9; | Length 24545; |
|-----------------------|---------------------------------------|---|-----------|-----------------|
| Best Local Similarity | 81.7%; | Pred. No. 6.3e-40; | | |
| Matches 255; | Conservative 0; | Mismatches 57; | Indels 0; | Gaps 0; |
| QY | 163 | CACAGAGCCCAAGGCGAGTGTTCGCAAGGAGAGCAGTGTGTCATATGAGGTGCGGACTGA | 222 | |
| Db | 10611 | CCGAGGCCCAAGTGGTGTTCCTTAGAGAGCAGCTGGCACGCGAAGAGCTGCAGGCAGA | 10670 | |
| QY | 223 | GGCAGGGGCCAGTGGCCACACTCAGCTGTGAGTGGCCCCAGGCCACAGAGGTGACGTG | 282 | |
| Db | 10671 | GGCAGGAGCCAGTGGCCACACTCAGCTGCAGGTGGCCCCAGGCCACAGAGGTGACGTG | 10730 | |
| QY | 283 | GTACAGAGTAGGGAAGAGAGCTCAGCTCCAGTTTCGAAGTGGCGATAGAGGGCTGCGGCTG | 342 | |
| Db | 10731 | GTACAAAGTAGGGAAGAGAGCTCAGCTCCAGTTTCGAAGTGTGCATGGAGGCCACAGGCTG | 10790 | |
| QY | 343 | CATCGGCGAGCTGTGTGTCACAGCGAGGCCAGGCAGATGCTGGGGAGTACACCTGTGA | 402 | |
| Db | 10791 | CACGCGCAGGCTGGTTGTGCAGCAGCGAGGCCAGCGGATGCGGGGAGTATAGCTGCGA | 10850 | |
| QY | 403 | GGCTGGGGCGACGCGCTCTCCCTCCACCTTGATCTTCAGAGCCCAAGGCGGTGTTCG | 462 | |
| Db | 10851 | GGCTGGGGCGCACGCGCTCTCCCTCCATCTGGATGTCAAAGTCAAGTGTGATAGAGCAGAC | 10910 | |
| QY | 463 | AAAGGAGCAGCT | 474 | |
| Db | 10911 | ATTTGAGCATCT | 10922 | |
| RESULT 5 | | | | |
| AX883644 | | | | |
| LOCUS | AX883644 | 2488 bp | DNA | linear |
| DEFINITION | Sequence 18549 from Patent EP1074617. | | | PAT 17-DEC-2003 |

| | |
|---|---|
| ACCESSION | AX883644 |
| VERSION | AX883644.1 |
| KEYWORDS | GI:40038545 |
| SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| AUTHORS | Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T. |
| TITLE | Primers for synthesizing full-length cDNA and their use |
| JOURNAL | Patent: EP 1074617-A 18549 07-FEB-2001; |
| FEATURES | Research Association for Biotechnology (JP) Location/Qualifiers 1..2488 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606" 697_..1197 /note="unnamed protein product" /codon_start=1 /protein_id="CAE92053.1" /gb_xref="GI:40038546" /translation="MALHGSMNCSCAGRWMGFSQLGHQCCOAWEGSVSPTFSSRPYS PVWYIDAVSLTSCVPPCPSEPKVPFAKEQAHREVOAEAGASATLSCVEVAQAQTE FWIYKDGLKLSRSSKVRVEAVGCTRLRVVQQAGQABEGSYSEAGGGQLSPRLQVAGQ CFGDAE" |
| ORIGIN | |
| Query Match | 45.1%; Score 218.6; DB 6; Length 2488; |
| Best Local Similarity | 86.1%; Pred. NO. 2.8e-39; |
| Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0; | |
| Qy | 163 CACAGAGCCCAAGCAGTCCTTGCCCAAGGAGCAGTTGTGTGCATAATGAGGTGGGACTGA 222 |
| Db | 891 CCAGAGCCCAAGTGGTGTGTGCCAAGGAGCAGCACACAGGAGGTGCAGGCTGA 950 |
| Qy | 223 GGCGGGGGCCAGTGCACACTGAGCTGTGAGGTGGCCCCAGGCCCACAGACAGATGACGCTG 282 |
| Db | 951 GGCGGGGGCCAGTGCACACTGAGCTGCAGGTGGCCCCAGGCCCACAGACAGATGACGCTG 1010 |
| Qy | 283 GTACAAGATGGAAAGACTGAGCTCCAATTGCAAGTGGCATAGAGCTGCGGGCTG 342 |
| Db | 1011 GTACAAGATGGAAAGAAGCTGAGTTCAGACTCGAAGTGCAGTGGAGCGCTGGGGCTG 1070 |
| Qy | 343 CATGCCGACGCTGGTGGTGCAGCAGGCGAGCCAGGCAGATGCTGGSGAGTACACCTGTGA 402 |
| Db | 1071 CACACGAGGCTGTGTGTGCAGCAGCGGGCCACAGCAGAGGCGGGAGTACAGCTGCGA 1130 |
| Qy | 403 GGCTGGGGGCCAGCGGCTCTCCCTCCACTCGATGTTTCAG 443 |
| Db | 1131 GGCAGGGGCTCAGCAGCTCTCCTTCGCGCTGCAGGTGGCAG 1171 |
| RESULT 6 | |
| BD160445 | 2488 bp DNA linear PAT 17-JAN-2003 |
| LOCUS | BD160445 |
| DEFINITION | Primer for synthesizing full-length cDNA and use thereof. |
| ACCESSION | BD160445 |
| VERSION | BD160445.1 |
| KEYWORDS | GI:27866203 |
| SOURCE | JP 2002191363-A/15288. |
| ORGANISM | Homo sapiens (human) |
| REFERENCE | Homo sapiens |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 2488) |
| TITLE | Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T. |
| JOURNAL | Primer for synthesizing full-length cDNA and use thereof Patent: JP 2002191363-A/15288 09-JUL-2002; |
| COMMENT | OS Homo sapiens (human) PN JP 2002191363-A/15288 |

09-JUL-2002
28-JUL-2000 JP 2002080990
TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
SAITO,
JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
KEIICHI NAGAI, TETSUJI OTSUKI

```

C12N15/09.C07K14/47.C07K16/18.C12N1/15.C12N1/19.C12N1/21.C12N5/
10.PC.C12P1/02.C12Q1/68/C12P2/08.G05F17/30.C12N15/00.C12N5/00.CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
(697)..(1194).
FT CDS Location/Qualifiers
1..2488
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

| | | | | |
|-----------------------|-----------------|----------------------------|--|--|
| Query Match | 45.1% | Score 218.6; | DB 6; | Length 2488; |
| Best Local Similarity | 86.1%; | Pred. No. 2.8e-39; | | |
| Matches 242; | Conservative 0; | Mismatches 39; | Indels 0; | Gaps 0; |
| QY | 163 | CACAGAGCCCAAGCAGCTGTTTGC | CAAGAGCAGATTGCTG | CAATATGAGGTGCGGACTTGA 222 |
| Db | 891 | CCACAGAGCCCAAGCTGGTGGT | TTTGC | CAAGGAGCAGCCAGCACACAGGGAGGTGCAGGCTGA 950 |
| QY | 223 | GCAGGGGCGCAGTGCACACTGAGCT | TGAGGTGGCCAGGCCACAGACAGAGGTTGACGTG 282 | |
| Db | 951 | GGCGGGGGCCAGTGCACAGCTGACGT | GCAGGTTGGCCCCAGGCCACAGACAGAGGTTGACGTG 1010 | |
| QY | 283 | GTACAAGGATGGGAAGAAGCTGAGCT | CACAGTTCGAAAGTGCGCATAGAGGCTGGCGGCTG 342 | |
| Db | 1011 | GTACAAGGATGGGAAGAAGCTGAGT | TTCAGCTCGAAAGTGCAGTGGAGGCGGTGGGCTG 1070 | |
| QY | 343 | CATGGGCGAGCTGGTGGTGAGCAGG | CAGGCGAGGACAGTGTGGGGAGTACACCTGTGA 402 | |
| Db | 1071 | CACCGGAGGCTGGTGGTGAGCAGG | CGGGCCAGGCAGAGGCGGGGAGTACAGCTCGCA 1130 | |
| QY | 403 | GGCTGGGGGCGCAGCGGGCTCTCCT | TTCACCTGGATGTTTCAG 443 | |
| Db | 1131 | GGCAGGGGGTTCAGCAGCTCTCTT | CGGCTTCGAGGTGGCAG 1171 | |

| | |
|------------|--|
| RESULT 7 | |
| AC024186 | |
| LOCUS | 2488 bp mRNA linear PRI 01-AUG-2002 |
| DEFINITION | Human sapiens CDNA FLJ141124 fis, clone MAMWA1002498. |
| ACCESSION | AK024186 |
| VERSION | AK024186.1 GI:10436503 |
| KEYWORDS | oligo capping; fis (full insert sequence). |
| SOURCE | Human sapiens (human) |
| ORGANISM | Human sapiens |
| | Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; |
| | Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. |
| REFERENCE | 1 |
| AUTHORS | Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Takahashi-Fujii, A., Hara, H., Tanase, T., Nomura, Y., Todiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saiko, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y. and Oshima, A. |
| | NEDO human cDNA sequencing project |
| TITLE | Unpublished |
| JOURNAL | 2 (bases 1 to 2488) |
| REFERENCE | Isogai, T. and Otsuki, T. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1530-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) |
| COMMENT | NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library |

| | | | |
|--|------------|---|-----------------|
| | AC023889 | linear | HTG 07-JUL-2000 |
| | LOCUS | Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT | |
| | DEFINITION | SEQUENCE, 13 unordered pieces. | |
| | ACCESSION | AC023889 | |
| | VERSION | AC023889.3 GI:8969253 | |
| | KEYWORDS | HTG; HTGS_PHASE1; HTGS_DRAFT. | |
| | SOURCE | Homo sapiens (human) | |
| | ORGANISM | Homo sapiens | |
| | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| | | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| | REFERENCE | 1 (bases 1 to 174612) | |
| | AUTHORS | Waterston,R.H. | |
| | TITLE | The sequence of Homo sapiens clone | |
| | JOURNAL | Unpublished | |
| | REFERENCE | 2 (bases 1 to 174612) | |
| | AUTHORS | Waterston,R.H. | |
| | TITLE | Direct Submission | |
| | JOURNAL | Submitted (18-FEB-2000) Genome Sequencing Center, Washington | |
| | | University School of Medicine, 4444 Forest Park Parkway, St. Louis, | |
| | | MO 63108, USA | |
| | COMMENT | On Jul 7, 2000 this sequence version replaced gi:8748947. | |

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

| | |
|-----------|---|
| REFERENCE | 1. Young, P., Ehler, E. and Gautel, M. |
| AUTHORS | Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly |
| TITLE | Unpublished |
| JOURNAL | 2 (bases 1 to 10091) |
| REFERENCE | Gautel, M.S. |
| AUTHORS | Direct Submission |
| TITLE | Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie, Max-Planck-Institut fuer Molekulare Physiologie, Otto-Rahn-Strasse 11, Dortmund, 44227, GERMANY |
| JOURNAL | 11, Dortmund, 44227, GERMANY |
| FEATURES | Location/Qualifiers |
| source | 1..10091 |
| | /organism="Homo sapiens" |
| | /mol_type="genomic DNA" |
| | /db_xref="taxon:9606" |
| gene | 1265..6155 |
| | /gene="OBSCN" |
| mRNA | Join(<1265..1540,5890..>6155) |
| | /gene="OBSCN" |

| | | |
|---|---|---|
| CDS | join(<1265..1540,5880..>6155) /gene="OBSCN" /codon_start=3 /product="obscurin" /protein_id="CAC85749.1" /db_xref="GI:21104332" /db_xref="SPTREMBL:Q8NH6" /translation="PKAVFAKEQPAHREVQAEAGASATLSCEVAQAQTEVTWYKDGKK LSSSLKVEAGATRRLLVVOAGQADTGEYCEAGGQQLSFLQVAEPRAVFAKQOP ASREVQAEAGTSATLSCEVAQAQTEVTWYKDGKKLSSSKVRMEAVGCTRRLLVVQAG QADAGEYCKAGDQRLSFLHVA" 1265..1540 /gene="OBSCN" /note="B" 1541..5879 /gene="OBSCN" /number=16 5880..>6155 /gene="OBSCN" /number=17 | Unpublished 2 (bases 1 to 3956) Isogai,T. and Yamamoto,J. Direct Submission Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail:genomics@kri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction; Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB. |
| exon | | |
| intron | | |
| exon | | |
| ORIGIN | | |
| Query Match 44.5%; Score 216; DB 9; Length 10091; Best Local Similarity 80.8%; Pred. No. 8.7e-39; Matches 252; Conservative 0; Mismatches 60; Indels 0; Gaps 0; | | |
| QY 155 CTGCACATCAGAGCCCAAGCGAGTGTTCACAGGAGCAGTGTGTCATATGAGGTG 214 | | |
| Db 1252 CTGCTTCCCCAGAGCCCAAGCGGTGTTCCTAAGAGCAGCAGCAGCAGGAGGTG 1311 | | |
| QY 215 CGGACTGAGCGAGGGGCCAGTGCACACTGAGCTGTGAGTGGCCAGGCCACAGACAG 274 | | |
| Db 1312 CAGGCTGAAGCAGGGGCTAGCGCCACACTGAGCTGTGAGTGGCCAGGCCACAGACAG 1371 | | |
| QY 275 GTGACGTGTACAGGATGGGAAAGCTGAGCTCCAGTTCGAAAGTGCGCATAGAGGCT 334 | | |
| Db 1372 GTGACTTGGTACAGGACGGGAGAGCTGAGCTCCAGTTCGAAAGTGTCATGTGGAGGCT 1431 | | |
| QY 335 GCGGGCTGCATCGCGCAGCTGGTGTGCAGCAGGCGCAGGCGAGTGTGCGGGAGTAC 394 | | |
| Db 1432 CGAGCTGTACAGGAGGCTGGTGTGCAGCAGGCGCAGGCGAGTGTGCGGGAGTAT 1491 | | |
| QY 395 ACCTGTGAGGCTGGGGCCAGCGGCTCTCTTCACCTGGATGTTTCAGAGCCCAAGGCG 454 | | |
| Db 1492 AGCTGTGAGGCTGGGGCCAGCAGCTCTCTTCGCGCTGCAAGTGGCAGGTCAGTGTTT 1551 | | |
| QY 455 GTGTTTGCAAG 466 | | |
| Db 1552 GGGGATGCTGAG 1563 | | |
| RESULT 10 | | |
| AK128447/c | | |
| LOCUS AK128447 3956 bp mRNA linear PRI 09-SEP-2003 | | |
| DEFINITION Homo sapiens cDNA FLJ46590 fis, clone THYMU304441. | | |
| ACCESSION AK128447 | | |
| VERSION AK128447.1 GI:34535823 | | |
| KEYWORDS oligo capping; fis (full insert sequence). | | |
| SOURCE Homo sapiens (human) | | |
| ORGANISM Homo sapiens | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| REFERENCE 1 | | |
| AUTHORS Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T., Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Sugiyama,T., Irie,K., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T. NEDO human cDNA sequencing project | | |
| TITLE | | |
| JOURNAL Unpublished | | |
| REFERENCE 2 (bases 1 to 3956) | | |
| AUTHORS Isogai,T. and Yamamoto,J. | | |
| TITLE Direct Submission | | |
| JOURNAL Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail:genomics@kri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction; Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB. | | |
| COMMENT | | |
| FEATURES | | |
| source | | |
| 1..3956 | | |
| /organism="Homo sapiens" | | |
| /mol_type="mRNA" | | |
| /db_xref="taxon:9606" | | |
| /clone="THYMU304441" | | |
| /tissue_type="thymus" | | |
| /clone_lib="THYMU3" | | |
| /note="cloning vector: pME18SFL3" | | |
| 1014..1415 | | |
| /note="unnamed protein product" | | |
| /codon_start=1 | | |
| /protein_id="BAC87445.1" | | |
| /db_xref="GI:34535824" | | |
| /translation="MSPSPGPPHSSAWHWP PPPPPAPCPAPAPWQTPPWALGWTRR DAGSEAPQTQPMQPSDSQAGLSIIKQCTVTHRAPPTTRAKGLRLHVGSTIIMSLG NRMRPLCLGLPGFPGPHLSPALPTTQGRQ" | | |
| ORIGIN | | |
| Query Match 43.5%; Score 211.2; DB 9; Length 3956; Best Local Similarity 83.3%; Pred. No. 1.2e-37; Matches 240; Conservative 0; Mismatches 48; Indels 0; Gaps 0; | | |
| QY 156 TGCACATCACAGAGCCCAAGCGAGTGTTCACAGGAGCAGTGTGTCATATGAGGTGC 215 | | |
| Db 1139 TGTCCATCCAGAGCCCAAGGTGTTTTCACAGAGCAGCCGCGATGCGGAGGTGC 1080 | | |
| QY 216 GGACTGAGCGAGGGGCCAGTGCACACTGAGCTGTGAGTGGCCAGGCCACAGACAGAG 275 | | |
| Db 1079 AGGCGGAGGCGGGGCCAGTGCACCGTGTGAGTGGCCAGGCCAGATGAGGAGG 1020 | | |
| QY 276 TGACGTGTACAGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCATAGAGGCTG 335 | | |
| Db 1019 TGACATGTGTACAGGACCGGGAAGAGTGTGAGCTCAAAAGTGGCATAGAGGCCA 960 | | |
| QY 336 CCGGCTGCATGCGGCGAGCTGGTGTGCAGCAGGCGCCAGGCGAGTGTGCGGGAGTACA 395 | | |
| Db 959 GCGGCTACACAGGAGGCTGTGTAGTCAGCAGGCGGGCCAGCGGATGCTGGGGAGTACA 900 | | |
| QY 396 CTTGTGAGGCTGGGGGCCAGCGGCTCTCTTCACCTGGATGTTTCAG 443 | | |
| Db 899 GCTGTGAGGCGGGGGCCAAACGGCTGTCTTCCGCTGCACGTGGCAG 852 | | |
| RESULT 11 | | |
| HS314903 | | |
| LOCUS HS314903 6400 bp DNA linear PRI 21-MAY-2002 | | |
| DEFINITION Homo sapiens partial OBSCN gene for obscurin, exons C-22. | | |
| ACCESSION AJ314903 | | |
| VERSION AJ314903.1 GI:21104333 | | |
| KEYWORDS OBSCN gene; obscurin. | | |
| SOURCE Homo sapiens (human) | | |
| ORGANISM Homo sapiens | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| REFERENCE 1 | | |
| AUTHORS Young,P., Ehler,E. and Gautel,M. | | |
| TITLE Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere | | |

assembly
Unpublished
2 (bases 1 to 6400)
Gautel, M.S.
Direct Submission
Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie,
Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Strasse
11, Dortmund, 44227, GERMANY
Location/Qualifiers
1. .6400
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
2201..6263
/gene="OBSCN"
join(<2201..2468,3519..3794,4017..4295,4425..4691,
4819..5088,5998..56263)
/gene="OBSCN"
join(<2201..2468,3519..3794,4017..4295,4425..4691,
4819..5088,5998..56263)
/gene="OBSCN"
/codon_start=1
/product="obscurin"
/protein_id="CAC85750.1"
/db_xref="GI:21104334"
/db_xref="SPTREML:OBHNS"
/translation="VVFAKDQPVHREVCBAAGTSTMLSCVAQAQTEVMYKDKKLS
SSSKRVEAVGCTRLVQVQAGQADAGEISCEAGGQRLSLHLVAEKVPVPAEQPAC
REVQAEASATLSCEVAQGMETVYKDKGKSSSKVMEASGYTRRLVQVQAGQ
DAGYSCEAGQRLSFRHLVAELEPQISERCPREPLVKEHEDIILTAFLATPSAAT
VTWKDGVIEIRSRKHETASOGDTHLTTHGAQVLDALYSRVGAEQFPVQVEV
AAKCRILLEPCVGLGTTVLACELSPACAEVWRCNLTOLRVGKRFQMVAEQVRL
TVLGRADAGEVYCESDDHTSAQLTVSPVVKPMGSLTVVABEGGATFCQVVS
PSDVAVWFRGALLQSEKFAISQSGASHSLTLDLVLEDAGQITVEAGASSAAL
RVREAPVFRKKLEPQVVEERSVTLVELTRAWPELRLNATAPAGKNVEIHAG
ARHLVLHNVGFADRGFGCETPDDKTQAKLTVE"
/gene="OBSCN"
/note="C"
2469..3518
/gene="OBSCN"
/number=17
3519..3794
/gene="OBSCN"
/number=18
3795..4016
/gene="OBSCN"
/number=18
4017..4295
/gene="OBSCN"
/number=19
4296..4424
/gene="OBSCN"
/number=19
4425..4691
/gene="OBSCN"
/number=20
4692..4818
/gene="OBSCN"
/number=20
4819..5088
/gene="OBSCN"
/number=21
5089..5997
/gene="OBSCN"
/number=21
5998..56263
/gene="OBSCN"
/number=22

ORIGIN
Query Match 43.2%; Score 209.6; DB 9; Length 6400;
Best Local Similarity 83.0%; Pred. No. 2.6e-37;

Matches 239; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 156 TGCATATCACAGAGCCCAAGCAGTGTTCACAGAGCAGTGTGTGATATGAGGTGC 215
|||
3507 TGTCCATCCAGAGCCCAAGTGGTGTTCACAGAGCAGCCGCATGACGGAGGTGC 3566
|||
QY 216 GAGCTGAGGCGAGGGCCAGTGCACACTGAGCTGTGAGTGGCCAGGCCACAGACAGAGG 275
|||
3567 AGCGGAGGCGGGGCCAGTGCACCGTGTGAGTGGCCAGGCCACAGATGGAGG 3626
|||
QY 276 TGACGTGTGTACAAAGATGGAAAGAGTGTGAGTGTCCAGTTCGAAAGTGGCATAGAGGTG 335
|||
3627 TGACATGTGTACAAAGACGGGAAGAGTGTGAGTGTCCAGTTCGAAAGTGGCATGAGAGGCA 3686
|||
QY 336 CGGGCTGTGATCGGCGAGCTGGTGTGACAGCAGGAGGCCAGCAGATGCTGGGGAGTACA 395
|||
3687 GCGGCTGTGATCGGCGAGCTGGTGTGACAGCAGGAGGCCAGCAGATGCTGGGGAGTACA 3746
|||
QY 396 CTTGTGAGGCTGGGGCCAGCGGCTCTCTTCCACCTGAGATTTTCAAG 443
|||
DB 3747 GCTGTGAGGCGGGGGCCCAACGGCTGCTCTCCGCGTGCACGTGGCAG 3794
|||
RESULT 12
AL645854/c
LOCUS
DEFINITION
211829 bp DNA linear ROD 05-APR-2002
Mouse DNA sequence from clone RP23-192P17 on chromosome 11,
complete sequence.
ACCESSION
AL645854
VERSION
AL645854.10 GI:20068636
KEYWORDS
HTG.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
1 Smith, M.
AUTHORS
Direct Submission
TITLE
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
COMMENT
On Apr 7, 2002 this sequence version replaced gi:19699553.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-192P17 is
from the RPCI-23 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6.
Location/Qualifiers
1. .211829
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-192P17"
/clone_lib="RPCI-23"
FEATURES
source
ORIGIN

Query Match 42.8%; Score 207.6; DB 10; Length 211829;
 Best Local Similarity 82.9%; Pred. No. 4.4e-37;
 Matches 237; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 165 CAGAGCCCAAGCAGTGTTCGACAGGACGAGTGTGTCATATAGAGTGGCGGCTGAGG 224
 Db 91363 CAGAGCCCAAGCAGTGTTCGACAGGACGAGGACGAGTGTGTCATATAGAGTGGCGGCTGAGG 91304

QY 225 CAGGSGCCAGTCCACACTGAGTGTGAGTGGCCCGCCAGCCAGACAGAGTGTGAGTGTGT 284
 Db 91303 CTGGGGCCAGTCCACACTGAGTGTGAGTGGCCCGCCAGCCAGACAGTGTGAGTGTGTGT 91244

QY 285 ACAAGATGGGGAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGCA 344
 Db 91243 TCAAGGCGGGAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGCT 91184

QY 345 TCGGCGAGTGTGTGTCAGCAGGACGAGGACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGCT 404
 Db 91183 CAGGAGGAGTGTGTGTCAGCAGGACGAGGACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGCT 91124

QY 405 CTGGGGGCGGAGGCGGTCTCTCTCTTCCACCTGGATGTTTCAGAGCCCAA 450
 Db 91123 CGGGGGCCAGAGTCTCTCTCTGCGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGCT 91078

RESULT 13

AC099089/c

LOCUS

Rattus norvegicus clone CH230-154E3, WORKING DRAFT SEQUENCE.

AC099089

AC099089.5 GI:30522701

HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus

1 (bases 1 to 260998)

AC099089

Munzy,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,

Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falla,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,

Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,B., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kowis,C., Kraft,C.T., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,

Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

Lorensheva,L., Loulisede,H., Lozada,R.J., Lu,X., Ma,J.,

Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,

Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,

Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,

Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,

Nankervis,C., Neal,D., Newton,S., Nguyen,N., Norris,S.,

Nwaokeleneh,O., Okwunonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,

Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,

Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., L.,

TITLE

JOURNAL

REFERENCE

AUTHORS

Worley,K.C.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Center: Genome Center

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GGLZ

Center clone name: CH230-154E3

Assembly program: Atlas 3.0;

Consensus quality: 216929 bases at least Q40

Consensus quality: 219228 bases at least Q30

Consensus quality: 220845 bases at least Q20

Estimated insert size: 226118; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
 consists of 1 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 260998: contig of 260998 bp in length.
 Location/Qualifiers
 1..260998
 /organism="Rattus norvegicus"

FEATURES

source

```

/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-154E3"
misc_feature 1..1486
note="wgs_contig"

ORIGIN
Query Match 42.8%; Score 207.4; DB 2; Length 260998;
Best Local Similarity 83.6%; Pred. No. 4.8e-37;
Matches 235; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 163 CACAGAGCCCAAGCAGTGTTCACAGGACGAGTGTGTCATATGAGGTGCGGACTGA 222
Db 188410 CCCAGAGCCCAAGTGTGTTCACAGGACGAGCAGCAGCAGTGTGAGGAGGAGA 188351

QY 223 GCGAGGGCCCAAGTGTTCACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 282
Db 188350 GCGAGGGCCCAAGTGTTCACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 188291

QY 293 GTACAGAGATGGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 342
Db 188290 GTTCAAGGACGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 188231

QY 343 CATGCGGAGTGTGTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402
Db 188230 CTCAGAGGAGTGTGTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 188171

QY 403 GCGTGGGGCCAGCGGCTCTCTCCACCTGAGTGTTCAG 443
Db 188170 GCGCGGGGACAGAGGCTCTCTCCGCTGAGCGTGGCAG 188130

RESULT 14
AC026657 164766 bp DNA linear HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 1 clone RP11-245P10, WORKING DRAFT
DEFINITION AC026657
AC026657 164766 bp DNA linear HTG 01-SEP-2000
VERSION AC026657.4 GI:9958202
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164766)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 164766)
Waterston,R.H.
Direct Submission
Submitted (22-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7637349.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0245P10
----- Summary Statistics -----
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 139864 bases at least Q40
Consensus quality: 147666 bases at least Q30
Consensus quality: 151469 bases at least Q20
Insert size: 169000; agarose-fp
Insert size: 161074; sum-of-contigs

```

FEATURES

Location/Qualifiers

Quality coverage: 3.60 in Q20 bases; agarose-fp
Quality coverage: 3.92 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1124: contig of 1124 bp in length
1125 1224: gap of unknown length
1225 1224: contig of 1808 bp in length
3033 3132: gap of unknown length
3132 3133: contig of 1361 bp in length
3133 4593: gap of unknown length
4593 4594: gap of unknown length
4594 5860: contig of 1267 bp in length
5860 5861: gap of unknown length
5861 5961: contig of 1710 bp in length
5961 7671: gap of unknown length
7671 7771: contig of 1899 bp in length
7771 9670: gap of unknown length
9670 9670: contig of 1946 bp in length
9670 11715: gap of unknown length
11715 11716: contig of 2428 bp in length
11716 14243: gap of unknown length
14243 14343: contig of 2544 bp in length
14343 16887: gap of unknown length
16887 16987: gap of unknown length
16987 19247: contig of 2260 bp in length
19247 19347: gap of unknown length
19347 21375: contig of 2028 bp in length
21375 21475: gap of unknown length
21475 25025: contig of 3550 bp in length
25025 25125: gap of unknown length
25125 28051: contig of 2926 bp in length
28051 28151: gap of unknown length
28151 32054: contig of 3903 bp in length
32054 32154: gap of unknown length
32154 35716: contig of 3562 bp in length
35716 35816: gap of unknown length
35816 40891: contig of 5075 bp in length
40891 40991: gap of unknown length
40991 44027: contig of 3036 bp in length
44027 44127: gap of unknown length
44127 48899: contig of 4772 bp in length
48899 48999: gap of unknown length
48999 54255: contig of 5256 bp in length
54255 54355: gap of unknown length
54355 58994: contig of 4639 bp in length
58994 59094: gap of unknown length
59094 64390: contig of 5296 bp in length
64390 64490: gap of unknown length
64490 70865: contig of 6375 bp in length
70865 70965: gap of unknown length
70965 78667: contig of 7702 bp in length
78667 78668: gap of unknown length
78668 87207: contig of 8440 bp in length
87207 87307: gap of unknown length
87307 97858: contig of 10551 bp in length
97858 97959: gap of unknown length
97959 109520: contig of 11562 bp in length
109520 109620: gap of unknown length
109620 120908: contig of 11288 bp in length
120908 121008: gap of unknown length
121008 141477: contig of 20469 bp in length
141477 141577: gap of unknown length
141577 162672: contig of 21095 bp in length
162672 162772: gap of unknown length
162772 163974: contig of 1202 bp in length
163974 164074: gap of unknown length
164074 164766: contig of 892 bp in length.

source

1. .164766
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-245P10"

misc_feature
1. .1124
/note="assembly_name:Contig14"

misc_feature
1225. .3032
/note="assembly_name:Contig14"

misc_feature
3133. .4493
/note="assembly_name:Contig23"

misc_feature
4594. .5860
/note="assembly_name:Contig24"

misc_feature
5961. .7670
/note="assembly_name:Contig25"

misc_feature
7771. .9669
/note="assembly_name:Contig26"

misc_feature
9770. .11715
/note="assembly_name:Contig27"

misc_feature
11816. .14243
/note="assembly_name:Contig28"

misc_feature
14344. .16887
/note="assembly_name:Contig29"

misc_feature
16988. .19247
/note="assembly_name:Contig30"

misc_feature
19348. .21375
/note="assembly_name:Contig31"

misc_feature
21476. .25025
/note="assembly_name:Contig32"

misc_feature
25126. .28051
/note="assembly_name:Contig33"

misc_feature
28152. .32054
/note="assembly_name:Contig34"

misc_feature
32155. .35716
/note="assembly_name:Contig35"

misc_feature
35817. .40891
/note="assembly_name:Contig36"

misc_feature
40992. .44027
/note="assembly_name:Contig37"

misc_feature
44128. .48899
/note="assembly_name:Contig38"

misc_feature
49000. .54255
/note="assembly_name:Contig39"

misc_feature
54356. .58994
/note="assembly_name:Contig40"

misc_feature
59095. .64390
/note="assembly_name:Contig41"

misc_feature
64491. .70865
/note="assembly_name:Contig42"

misc_feature
70966. .78667
/note="assembly_name:Contig43"

misc_feature
78768. .87207
/note="assembly_name:Contig44"

misc_feature
87308. .97858
/note="assembly_name:Contig45"

misc_feature
97959. .109520
/note="assembly_name:Contig46"

misc_feature
109621. .120908
/note="assembly_name:Contig47"

misc_feature
121009. .141477
/note="assembly_name:Contig48"

misc_feature
141578. .162672
/note="assembly_name:Contig49"

misc_feature
162773. .163974
/note="assembly_name:Contig50"

misc_feature
164075. .164766
/note="assembly_name:Contig51"

174 AGGCAGTGTGTTGCCAAGGACGAGTTGGTCATATAGAGTGCGGACTGAGGCAAGGGCCCA 233
|||||
54804 AGGTGTGTGTTGCCAAGGATCAGCCGGTGCACAGGGAAGTGCAGGCTGAGGCAAGGACCA 54866

234 GTGCCACACTGAGCTGTGAGTGTGCCAGGCCAGCCAGACAGAGGTGACGTGTACAAGATG 293
|||||
54864 GCACCATCTGAGCTGTGAGTGTGCCCAAGCCAGACGGAGGTATGTGTACAAGGACG 54923

294 GGAGAAGCTGAGCTCCAGTTTCGAAAGTGGCGCATAGAGGCTGCGGGCTCATCGCGCACG 353
|||||
54924 GGAAGAAGCTGAGCTCCAGTTTCGAAAATGCGTGTGGAGGCCGTGTGCACACGAGGAC 54983

354 TGGTGGTGCACAGGACGAGCCAGCAGATCTGGGAGTAGACCTGTGAGGCTGAGGCTGGGGCC 413
|||||
54984 TGTTGTGTCAGAGGACGAGCCAGCGGACCGCCGGGAGTACAGCTCGAGGCTGGGGGCC 55043

414 ACGCGCTCTCTTCCACTCGATGTTTCAG 443
|||||
55044 ACGCGCTCTCTTCCACTCTGCATGTGGCTG 55073

RESULT 15
AC142478/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-457UJ14, *** SEQUENCING IN PROGRESS
***, 25 unordered pieces.
AC142478
AC142478.1 GI:29374186
HTG: HTGS PHASE1.
VERSION
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 44358)
Muzny,D.,Marie,, Metzker,M.Lee,, Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,
Anyalebechi,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,B., Burrell,K., Calderon,E.,
Cardenas-V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,L., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,W.,
Gebregeorgis,E., Geert,K., Gill,R., Grady,M., Guerra,W., Guvera,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levay,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Loulseghe,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,K., Mallory,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mathew,S., McLeod,M., McNeill,T., Meenen,E., Miloslavjevic,A.,
Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
Olatunmbatan,A., Pal,S., Parks,K., Paetznick,S., Paul,H.,
Olarnunsaenon,A.

ORIGIN

| | | | | |
|-----------------------|--------------|--------------------|----------------|-------------------|
| Query Match | 41.5%; | Score 201.2; | DB 2; | Length 164766; |
| Best Local Similarity | 84.1%; | Pred. No. 1.3e-35; | | |
| Matches 227; | Conservative | 0; | Mismatches 43; | Indels 0; Gaps 0; |

| | 1 | 1402: | contig | of 1402 | bp in length |
|---|-------|--------|--------|------------|--------------|
| * | 1403 | 1502: | gap | of unknown | length |
| * | 1503 | 2892: | contig | of 1390 | bp in length |
| * | 2893 | 2992: | gap | of unknown | length |
| * | 2993 | 4166: | contig | of 1174 | bp in length |
| * | 4167 | 4266: | gap | of unknown | length |
| * | 4267 | 5505: | contig | of 1239 | bp in length |
| * | 5506 | 5605: | gap | of unknown | length |
| * | 5606 | 6651: | contig | of 1046 | bp in length |
| * | 6652 | 6751: | gap | of unknown | length |
| * | 6752 | 7800: | contig | of 1049 | bp in length |
| * | 7801 | 7900: | gap | of unknown | length |
| * | 7901 | 9334: | contig | of 1434 | bp in length |
| * | 9335 | 9434: | gap | of unknown | length |
| * | 9435 | 10944: | contig | of 1510 | bp in length |
| * | 10945 | 11044: | gap | of unknown | length |
| * | 11045 | 12272: | contig | of 1228 | bp in length |
| * | 12273 | 12372: | gap | of unknown | length |
| * | 12373 | 13628: | contig | of 1256 | bp in length |
| * | 13629 | 13728: | gap | of unknown | length |
| * | 13729 | 15016: | contig | of 1288 | bp in length |
| * | 15017 | 15116: | gap | of unknown | length |
| * | 15117 | 16392: | contig | of 1276 | bp in length |
| * | 16393 | 16492: | gap | of unknown | length |
| * | 16493 | 18284: | contig | of 1792 | bp in length |
| * | 18285 | 19384: | gap | of unknown | length |

Search completed: September 19, 2004, 19:27:03
Job time : 1982.42 secs

10/045,072.
1533.079002

412

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:07:04 ; Search time 1122.53 Seconds
(without alignments)
10242.013 Million cell updates/sec

Title: US-10-077-130-4_COPY_10286_10670

Perfect score: 385

Sequence: 1 acatagcctgaggcaggacg.....cctcagccacgctcaccatc 385

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55025578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: gb_est6.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|--------------------|
| 1 | 285.8 | 74.2 | 595 | 10 | BE086827 | BE086827 QV1-BT067 |
| 2 | 263.4 | 68.4 | 486 | 12 | BM030364 | BM030364 488988 MA |
| 3 | 252.6 | 65.6 | 536 | 10 | BF824937 | BF824937 ILO-HN003 |
| 4 | 240 | 62.3 | 723 | 13 | BQ446463 | BQ446463 UI-H-EU1- |

| | | | | | | | |
|---|----|-------|------|------|----|----------|---------------------|
| C | 5 | 239.4 | 62.2 | 487 | 10 | BE466441 | BE466441 hz21h03.x |
| C | 6 | 189 | 49.1 | 570 | 29 | CE836029 | CE836029 tigr-gss- |
| C | 7 | 186.2 | 48.4 | 437 | 10 | BF826423 | BF826423 CM4-HN002 |
| C | 8 | 175.2 | 45.5 | 359 | 10 | BF853346 | BF853346 MR2-EN009 |
| C | 9 | 168.8 | 43.8 | 369 | 10 | BF826489 | BF826489 CM4-HN002 |
| C | 10 | 165.4 | 43.0 | 324 | 13 | BQ361458 | BQ361458 PM0-OT023 |
| C | 11 | 163.2 | 42.4 | 601 | 10 | BF398067 | BF398067 UI-R-B52- |
| C | 12 | 158.8 | 41.2 | 690 | 28 | BH025901 | BH025901 RPT-24-2 |
| C | 13 | 127.2 | 33.0 | 505 | 29 | CE078765 | CE078765 tigr-gss- |
| C | 14 | 122.4 | 31.8 | 204 | 12 | BG987697 | BG987697 PM3-HT116 |
| C | 15 | 121.8 | 31.6 | 647 | 28 | AZ590254 | AZ590254 1M0399B18 |
| C | 16 | 112.4 | 29.2 | 247 | 12 | BG987696 | BG987696 PM3-HT116 |
| C | 17 | 107.8 | 28.0 | 628 | 28 | AZ985514 | AZ985514 2M0267G22 |
| C | 18 | 106.2 | 27.6 | 657 | 28 | AZ508477 | AZ508477 1M0347H09 |
| C | 19 | 103 | 26.8 | 350 | 10 | BF748073 | BF748073 MR2-BN038 |
| C | 20 | 86.2 | 22.4 | 742 | 13 | BJ358283 | BJ358283 603477905 |
| C | 21 | 84.8 | 22.0 | 380 | 10 | BF957683 | BF957683 PM1-NN120 |
| C | 22 | 82.2 | 21.4 | 382 | 9 | AI478779 | AI478779 tm24e02.x |
| C | 23 | 82.2 | 21.4 | 740 | 13 | BU689712 | BU689712 UI-CF-FN0 |
| C | 24 | 78.2 | 20.3 | 212 | 10 | BF827521 | BF827521 CM4-HN002 |
| C | 25 | 70 | 18.2 | 510 | 13 | EX478907 | EX478907 DKZ2P686N |
| C | 26 | 69 | 17.9 | 330 | 28 | AQ237886 | AQ237886 RPT111-68 |
| C | 27 | 64 | 16.6 | 173 | 10 | BE086762 | BE086762 QV1-BT067 |
| C | 28 | 57.8 | 15.0 | 586 | 29 | EX153687 | EX153687 Danilo rer |
| C | 29 | 54 | 14.0 | 332 | 28 | AZ754571 | AZ754571 cq01e04.f |
| C | 30 | 52 | 13.5 | 128 | 12 | BM031705 | BM031705 497184 MA |
| C | 31 | 52 | 13.5 | 997 | 29 | CNS05GFS | AL336218 Tetraodon |
| C | 32 | 50.4 | 13.1 | 722 | 13 | EU322682 | EU322682 603851547 |
| C | 33 | 49.4 | 12.8 | 1010 | 29 | CNS02HPP | AL197926 Tetraodon |
| C | 34 | 48.4 | 12.6 | 1036 | 29 | CNS00DAP | AL067111 Drosophil |
| C | 35 | 47 | 12.2 | 1173 | 28 | CC218835 | CC218835 CH261-69E |
| C | 36 | 46.6 | 12.1 | 1000 | 13 | EX407619 | EX407619 BX407619 |
| C | 37 | 46.6 | 12.1 | 1970 | 29 | CG748837 | CG748837 P042-4-E0 |
| C | 38 | 45.8 | 11.9 | 828 | 13 | EX425797 | EX425797 BX425797 |
| C | 39 | 45.6 | 11.8 | 855 | 28 | BZ637045 | BZ637045 OGCAN08TC |
| C | 40 | 45.2 | 11.7 | 354 | 10 | BF353892 | BF353892 ILS-HT070 |
| C | 41 | 45 | 11.7 | 871 | 29 | CNS04FAX | AL288114 Tetraodon |
| C | 42 | 44.8 | 11.6 | 658 | 13 | EX087878 | EX087878 BX087878 |
| C | 43 | 44.6 | 11.6 | 1181 | 12 | BT869996 | BT869996 603394150 |
| C | 44 | 44.4 | 11.5 | 1101 | 29 | CNS0153P | AL104949 Drosophil |
| C | 45 | 44.4 | 11.5 | 1101 | 29 | CNS016JS | AL106634 Drosophil |

ALIGNMENTS

| | | | | | |
|------------|---|------------|--------------|--------|------------------|
| RESULT 1 | BE086827 | 555 bp | mRNA | linear | EST 12-JUN-2000 |
| LOCUS | QV1-BT0678-300400-182-f06 | BT0678 | Homo sapiens | cdna | mRNA sequence. |
| DEFINITION | BE086827 | | | | |
| ACCESSION | BE086827.1 | GI:8477221 | | | |
| VERSION | EST. | | | | |
| KEYWORDS | Homo sapiens (human) | | | | |
| SOURCE | Homo sapiens | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 595) | | | | |
| AUTHORS | Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. | | | | |
| TITLE | Shotgun sequencing of the human transcriptome with ORF expressed sequence tags | | | | |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. | 97 (7) | | | 3491-3496 (2000) |
| MEDLINE | 20202683 | | | | |
| PubMed | 10737800 | | | | |
| COMMENT | Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, | | | | |

[illegible]

plasmids were derived from ORS155 PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) and inserted into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

| ORIGIN | low stringency conditions." | | | |
|-----------------------|-----------------------------|---|--------|-------------------|
| Query Match | 48.4%; | Score 186.2; | DB 10; | Length 437; |
| Best Local Similarity | 78.0%; | Prod. No. 1.7e-30; | | |
| Matches | 224; | Conservative | 63; | Indels 0; Gaps 0; |
| QY | 99 | GGACCTCTGCCACTCTCACCGGTGAGGCCCTGCCAGCCAGATTCACAGAGGCTCAGGA | 158 | |
| DB | 425 | GTTCCTCAGCTACATCTATGTCAGGGCCCTGCTTGCAGACTCTATAGAGATATAGAA | 366 | |
| QY | 159 | ATGAAGAGGCCGTGGAAAGGGCCACAGCCATCTTTGTGTGTGAACTGAGCAAGTGGCCC | 218 | |
| DB | 365 | ACCAGAAGGCCACAGAAGGGGTACAGTCACATTTGCAATGTAAAGTGAAGAGCGCGCCC | 306 | |
| QY | 219 | CTGTGAGATGGAGAAAGGGGCCGAGAACTTCAGAGATGGGGACAGATACATCTCAGGC | 278 | |
| DB | 305 | CCGTGAGTGGAGAAAGGGGCCCAACCCCTCAAAGATGGGACAGTACAGCCTGAAGC | 246 | |
| QY | 279 | AGAGGGGCCACAGGTGTGAGCTGAGATCTGTGCCTTGGCCATGGCGGACGCCGGGAGT | 338 | |
| DB | 245 | AGATGGGACCACTGTGTAGCTTGCAATCTGTGCCTTGGTCATGACAGATCTCTGGAGAT | 186 | |
| QY | 339 | ACTGTGTGTGTGTCGGGCGAGGAGGACCTCAGCCACGCTCAACATC | 385 | |
| DB | 185 | ACTGTGTGCATATGTGACAGAGAGAGACCTCGGGCAGCGCCATGTG | 139 | |

339 ACTTGTGTGTGCGGAGGAGGACCTCAGCAGCGCTCACATC 385
 245 AGGATGGGACCACTGTGTAGCTTGAGATCTGTGGCTTGGTCATAGCAGATCTCTGGAGAA 186
 185 ACTGTGTGATATGTAGCAGGAGAGGACCTCGGCAGCGCCATGTG 139

[illegible]

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V.,
 O'Hare, M.J., Soares, F., Brenkani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

| | |
|---------|--|
| TITLE | Shotgun sequencing of the human transcriptome with ORF expressed sequence tags |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) |
| MEDLINE | 20202663 |
| PUBMED | 10737800 |
| COMMENT | Contact: Simpson A.J.G. |

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPSP/LICK Human Cancer Genome

Project. This entry can be seen in the following URL
<http://www.ludwig.org.br/scripts/gethtml2.pl?i=MR2&t2=MR2-EN0091-191200-004-f08&t3=2000-12-19&t4=1>
Seq primer: puc 18 forward
High quality sequence start: 19
High quality sequence stop: 158.

```

FEATURES
  source
    Location/Qualifiers
      1..359
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /dev_stage="Adult"
        /clone_lib="EN0091"
        /note="Organ: lung normal; Vector: puc18; Site 1: SmaI;
        Site 2: SmaI; A mini-library was made by cloning products
        derived from ORESTES PCR (U.S. Letters Patent application
        No. 196,716 - Ludwig Institute for Cancer Research)
        profiles into the pUC 18 vector. Reverse transcription of
        tissue mRNA and cDNA amplification were performed under
        low stringency conditions."
      ORIGIN
        Query Match      45.5%; Score 175.2; DB 10; Length 359;
        Best Local Similarity 83.5%; Pred. No. 3.9e-28;
        Matches 198; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
        QY 149 GGTCTGAGATGAAGAGCCCTGGAAGGGGCCACAGCCATGTTGTGTGTAAGTAC 208
        Db 359 GATGTGAGAAATACAGAGGCCACAGAGGGGCCACAGCTGTGCTGCAGTGTGAGCTGAGC 300
        QY 209 AAGTGGGCCCCCTGTGAGTGGAGGAGGGGCCGAGACCTCAGAGATGGGGACAGATAC 268
        Db 299 AAGNCGGCCCTGTGAGTGGAGGAGGGGCTCAGACCTCAGAGATGGGGACAGATAT 240
        QY 269 ATCTTGAGGAGGAGGAGGAGGAGTGTGAGTGCAGATCTGTGCGCTGGCCATGGCGGAC 328
        Db 239 AGCCTGAGGAGGAGGAGGAGGAGTGTGAGTGCAGATCTGTGCGCTGGCTGTGGAGGAC 180
        QY 329 GCGGGGAGTACTGTGTGTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 385
        Db 179 ACTGAGAGTATTGTGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 123
      RESULT 9
      BF826489
      LOCUS
      CM4-HN0020-221100-452-b09 HN0020 Homo sapiens cDNA, mRNA sequence.
      BF826489
      ACCESSION
      BF826489.1 GI:12169746
      VERSION
      EST.
      KEYWORDS
      SOURCE
      Homo sapiens (human)
      ORGANISM
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
      1 (bases 1 to 369)
      Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
      Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.,
      Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
      Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,
      O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
      Simpson, A.J.G.
      Shotgun sequencing of the human transcriptome with ORF expressed
      sequence tags
      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
      20202663
      10737800
      Contact: Simpson A.J.G.
      Laboratory of Cancer Genetics
      Ludwig Institute for Cancer Research
      Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
      Brazil
      Tel: +55-11-2704922
      Fax: +55-11-2707001
      Email: asimpson@ludwig.org.br
      This sequence was derived from the FAPESP/LICR Human Cancer Genome
      Project. This entry can be seen in the following URL
      (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-HN0020-
      221100-452-b09&t3=2000-11-22&t4=1)
      Seq primer: puc 18 forward
    REFERENCE
      AUTHORS
      TITLE
      JOURNAL
      MEDLINE
      PUBMED
      COMMENT
      Shotgun sequencing of the human transcriptome with ORF expressed
      sequence tags
      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
      20202663
      10737800
      Contact: Simpson A.J.G.
      Laboratory of Cancer Genetics
      Ludwig Institute for Cancer Research
      Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
      Brazil
      Tel: +55-11-2704922
      Fax: +55-11-2707001
      Email: asimpson@ludwig.org.br
      This sequence was derived from the FAPESP/LICR Human Cancer Genome
      Project. This entry can be seen in the following URL
      (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&t2=CM4-HN0020-
      221100-452-b09&t3=2000-11-22&t4=1)
      Seq primer: puc 18 forward
  
```


Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PMO&t2=PMO-070232-240501-005-a12&t3=2001-05-24&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 21
High quality sequence stop: 324.

FEATURES

Location/Qualifiers
1..324
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="GT0232"
/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 43.0%; Score 165.4; DB 13; Length 324;
Best Local Similarity 87.4%; Pred. No. 5.1e-26;
Matches 181; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 179 GCCACAGCAGCTTTGTGTGAACCTGAGCAAGTGGCCCTGTGCACTGAGGAGGGG 238
Db 324 GCCACGGCCAGCTCGGTGTGAGCTGAGCAAGGCGCCCTGTGAGTGGAGTGGAGAAAGGG 265
QY 239 CCCGAGACCTCAGAGATGGGACAGATACATCTCTGAGGACAGGAGGACCAAGTGTGAG 298
Db 264 TCCGAGACCTCAGAGATGGGACAGATACCTCTGAGGACAGGACGCGGCCATGTGTGAG 205
QY 299 CTGCAAGATCTGTGGCTGCCATGCGGACGCGCGGGAGTACTTGTGTGTGGGGCAG 358
Db 204 CTGCAAGATCCGTGGCTGGCCATGCTGTGATGCGCGGAGTACTCTGTGTGTGGAGAG 145
QY 359 GAGAGGACCTCAGCAGCTCACCATC 385
Db 144 GAGAGACCTCAGCCTTACTCACCATC 118

RESULT 11

BF398067/c
LOCUS
DEFINITION
UI-R-BS2-beg-h-06-0-UI.s1 UI-R-BS2 Rattus norvegicus cDNA clone
UI-R-BS2-beg-h-06-0-UI.3', mRNA sequence.

ACCESSION

BF398067
VERSION
KEYWORDS
SOURCE
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 601)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL

MEDLINE
PUBMED
COMMENT
Genome Res. 6 (9), 791-806 (1996)
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized embryo at 13 dpc library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers
1..601
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BS2-beg-h-06-0-UI"
/dev_stage="embryonic 13 dpc"
/lab_host="PH108 (Life Technologies)"
/clone_lib="UI-R-BS2"
/note="Vector: pT7P3D-Pac (pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BS2
library is a subtracted library derived from 13 dpc whole
embryo tissue. For a detailed description of the library
from which this clone was derived, please visit our web
site at ratstest.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG TISSUE=embryo at 13 dpc
TAG LIB=UI-R-BS2
TAG_SEQ=AATCC"

ORIGIN

Query Match 42.4%; Score 163.2; DB 10; Length 601;
Best Local Similarity 69.4%; Pred. No. 2.1e-25;
Matches 222; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 66 CTGGGAGTACTCTGTGTGTGGGAGGAGAGGACCTCTGCCACTCTCACCGTGAAGG 125
Db 517 CTAGGGCGTCCATGTCTCAGAGGGTGTGATTGTGTCTGTCTTCTTCTGACCTCCCGAG 458
QY 126 CCCTGCGCAGCAAGTTCACAGAGGGTCTGAGGAATGAGAGCGCGTGGAGGGCCACAG 185
Db 457 CCCTTCCGGCTTAAGTTCATAGAAAGTCTGAGAAATGAGGGGCCACAGAGGAACACAG 398
QY 186 CCATGTTGTGTGTAACCTGAGCAAGTGGCCCTCTGGAGTGGAGAGGGGCCCGAGA 245
Db 397 CCAGCTGAGCTGCAAACTGAGCAAGCGCTCGGTGACGTGGAAGAGGAGCAAGA 338
QY 246 ACCTCAGAGATGGGAGCAGATACATCTGAGGAGGAGGGGACAGGTGTGAGCTGCAGA 305
Db 337 CCTTGGAGACGGAGACAAATATGGCTGAGGACGACGAGCTGTGTGTGAGCTGCAGA 278
QY 306 TCTGTGGCTTGGCCATGCGGAGCGCCGGGAGTACTTGTGTGTGCGGCGAGAGAGGA 365
Db 277 TCCGTGGCTTACACAGCGCATCTGGGGAGTACTCATGTGTGTGGCGAGGAGA 218
QY 366 CCTCAGCAGCCTCACCATC 385
Db 217 CATCAGTGTCTGACTGTC 198

RESULT 12

BH025901/c
LOCUS
DEFINITION
RPCI-24-259H14.TV RPCI-24 Mus musculus genomic clone
RPCI-24-259H14, genomic survey sequence.

ACCESSION

VERSION
KEYWORDS
SOURCE
ORGANISM
BH025901
GI:14789365
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE
AUTHORS      Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
              Tsengye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
              Russell,D., de Jong,P. and Fraser,C.M.
TITLE        Mouse BAC End Sequences from Library RPCI-24
JOURNAL      Unpublished (1999)
COMMENT      Other GSSs: RPCI-24-259H14.TJ
              Contact: Shaying Zhao
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: szhao@tigr.org
              Clones are derived from the mouse BAC library RPCI-24. For BAC
              library availability, please contact Pieter de Jong
              (pdejong@mail.cho.org). Clones may be purchased from BACPAC
              Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end
              plate: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
              Plate: 259 row: H column: 14
              Seq primer: T7
              Class: BAC ends.

FEATURES
source      Location/Qualifiers
            1..690
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-24-259H14"
            /sex="Male"
            /cell_type="Spleen/Brain"
            /clone_lib="RPCI-24"
            /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI.
            RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
            library was cloned in the pTARBAC1 cloning vector at the
            BamHI sites using MboI partially digested male C57BL/6J
            DNA."

ORIGIN
Query Match      41.2%; Score 158.8; DB 28; Length 690;
Best Local Similarity 79.1%; Pred. No. 2e-24;
Matches 201; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 125 GCCCTGCCAGCCAGTTCACAGAGGGTCTGAGGATGAGAGGCGCTGGAAGGGGCCACA 184
DB 253 GCCCTGCCCTCCAGATTTCATGAGAGATTGAGAGGTCAGAGGCCACAGAGGCCCATG 194

QY 185 GCCATTGTGTGTGTAAGTCAAGAGTGGCCCTCTGGAGTGGAGAGGGGCCCGAG 244
DB 193 GCCACTCTGAGTGGCCAGATGAGCAGACTGCCCTCTGGAGTGGAGAGGGTCTGAG 134

QY 245 AACCTCAGAGTGGGACAGATACATCTCAGGACAGGGGACCAGGTGTGAGCTGCAG 304
DB 133 ACCCTGAGAGATGGGGCAGATACAGCTCAGGACGATGGCCCTGTGTGTGAGCTGCAG 74

QY 305 ATCTGTGCGCTGGCCATGGCCGAGCGCGGAGTACTTGTGTGTGTCGGGACGAGAGG 364
DB 73 ATCTGTGACCTGGTGTGTCGAGAGAGCTGGGAGTACTCATGTGTGTG-GGGCAGGAG 15

QY 365 ACCTCAGCCAGCT 378
DB 14 ACCTCAGCCACACT 1

RESULT 13
CE078765 LOCUS      CE078765      505 bp      DNA      linear      GSS 24-SEP-2003
DEFINITION      tigr-gss-dog-17000324480165 Dog Library Canis familiaris genomic,
                  genomic survey sequence.
ACCESSION      CE078765
VERSION        CE078765.1 GI:35145465
KEYWORDS       GSS.
SOURCE         Canis familiaris (dog)

ORGANISM      Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE      1 (bases 1 to 505)
AUTHORS        Kirkness,E.F., Rafna,V., Halpern,A.L., Levy,S., Remington,K.,
              Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
              Venter,J.C.
TITLE          The dog genome: survey sequencing and comparative analysis
JOURNAL        Science 301 (5641), 1898-1903 (2003)
MEDLINE        22875432
PUBMED         14512627
COMMENT        Contact: Kirkness EF
              The Institute for Genomic Research
              Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
              Rockville, MD 20850, USA
              Tel: 301-838-0200
              Fax: 301-838-0208
              Email: ekirknes@tigr.org
              Class: shotgun.
              Location/Qualifiers
              1..505
              /organism="Canis familiaris"
              /mol_type="genomic DNA"
              /strain="Standard Poodle"
              /db_xref="taxon:9615"
              /clone_lib="Dog Library"
              /note="Site 1: BstXI; Libraries were prepared from
              peripheral blood"

ORIGIN
Query Match      33.0%; Score 127.2; DB 29; Length 505;
Best Local Similarity 83.7%; Pred. No. 1.4e-17;
Matches 144; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 214 GGGCCCTGTGGAGTGGAGAGGGCCGAGAACCTCAGAGATGGGACAGATACATCCT 273
DB 2 GGGCCCTGTGGAGTGGAGAGGGCCCTCAGAGCCCTCAGAGCCGGGACAGGTCAGCCT 61

QY 274 GAGGACAGAGGGGACAGGTGTGAGCTGACATCTGCTGCTGCGCATGGCGGACGCCGG 333
DB 62 GAGGACAGAGGGGCGGTGTGCGAGCTGCGAGATCCGGGCTTGGCCATGGAGGATCCGG 121

QY 334 GGAGTACTTGTGTGTGTGCGGCGAGGAGGACCTTCAGCCACGCTCACCATC 385
DB 122 GGAGTACTTGTGTGTGTGCGGAGGAGGAGCGGCCGCTGCTCACCGTC 173

RESULT 14
BG987697/c LOCUS      BG987697      204 bp      mRNA      linear      EST 13-JUN-2001
DEFINITION      PM3-HT1165-110101-001-e07 HT1165 Homo sapiens cDNA, mRNA sequence.
ACCESSION      BG987697
VERSION        BG987697.1 GI:14391767
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 204)
AUTHORS        Dias Neto,E., Garcia Correa,R., Vertovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.P.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
              Brunstein,A., Geoliveira,P.S., Bucher,P., Jongeneel,C.V.,
              O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
TITLE          Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE        20202663
PUBMED         10737800
COMMENT        Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research

```

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpcsn@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtm2.pl?t1=PM3&t2=PM3-H7L165-110101-001-e07&t3=2001-01-11&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 35
High quality sequence stop: 204.

FEATURES

```

1. 204
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref=taxon:9606"
/dev_stage="Adult"
/Clone.lib="HT1165"
/notes=Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from CRESTS PCR (U.S. Letters Patent application
No. 1,936,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

ORIGIN

| | | | | |
|-----------------------|-----------------|--|-----------|-------------|
| Query Match | 31.8%; | Score 122.4; | DB 12; | Length 204; |
| Best Local Similarity | 75.0%; | Pred. No. 1e-16; | | |
| Matches 153; | Conservative 0; | Mismatches 51; | Indels 0; | Gaps 0; |
| QY | 7 | CCTGAGGCAGACGGGGCTGTGTGCGAGCTGCGAGATCTGTGGCCCTGGCTGTGGCAGATGC | 66 | |
| DB | 204 | CCTGAGGCAGACGGGACGAGGTGTGAGCTGCGATTCTGTGGCCTGGCTGTGGAGGCAC | 145 | |
| QY | 67 | TGGGAGGTACTCTGTGTGTGTGTGGGAGGAGAGGAACTCTGCCACTCTCACCGTGAAGGC | 126 | |
| DB | 144 | TGGAGAGTATTGTGTGTGTGTGTCGGGCAGGAGAGAACTCAGCTACACTCACTTTTCAGGCG | 85 | |
| QY | 127 | CTGTGCCAGCCAAAGTTTCACAGAGGGTCTGTAGGAATGAAGAGCGCCGTGGGAAGGGGCCACAGC | 186 | |
| DB | 84 | CTGTGCCAGCAGATTTCATACCCACATGACCNACCGAAGCCAGAGAGGGGCCACGCG | 25 | |
| QY | 187 | CATGTTGTGGTGTGAACCTGAGCAA | 210 | |
| DB | 24 | CACACTGCACTGTGAACCTGAGCAA | 1 | |

RESULT 15

| Accession | LOCUS | Accession | Size | Library | Species | Accession |
|------------|------------|---------------------|-------------------|--------------|-----------------|-----------|
| AZ590254/c | AZ590254 | 647 bp | DNA | linear | GSS 13-DEC-2000 | |
| | DEFINITION | 10kb plasmid | UUCGIC library | Mus musculus | genomic | |
| | | clone UUCGICM039818 | R. genomic survey | | sequence. | |

| | |
|-----------|----------|
| ACCESSION | AZ590254 |
| WEBCON | A7E90254 |

VERSION
KEYWORDS
A2390234.1
G1:11/12444
GSS

KEYWORDS
SOURCE

SOURCE Mus musculus
ORGANISM Mus musculus

NOT AFFRANCO

REFERENCE
1 (bases 1 to 647)

AUTHORS

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
Niederhausen, A. and Wright, D. Weiss, R.

TITLE

plasmid inserts

JOURNAL

COMMENT
Contact: Robert B. Weiss

University of Utah

84112 USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: cdunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0399 row: B column: 18
Seq primer: CACACGAAACAGCTGTACG
Class: plasmid ends
High quality sequence stop: 647.

FEATURES

source

```
1. .647
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0399B18"
/sex="Male"
```

sex-mate" E. Coli strain XL10-Gold, Tl-resistant, F-"/
/ad-host" Mouse 10kb plasmid UUGC1M library"
/clone.lib="Mouse 10kb plasmid UUGC1M library"
/note="vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA
was hydrothermally sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 [gi14732114|cb|AF129072.1], a copy-number
inducible derivative of plasmid R1, the vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN: "

Query Match 31.6%; Score 121.8; DB 28; Length 647;
Best Local Similarity 79.6%; Pred. No. 2.3e-16;
Matches 144; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| Qy | 125 | GCCTGCCAGCCCAAGTTTCACAGAGGGTCTCAGAGNATGAAGAGGCCGCTGGAGAGGGGCCACA | 181 |
| Db | 181 | GGCTTGCTGCCAAATTTATAGAAGATCTGAGAAGCCCAAGAAGCCACAGAAGGGGGCCACA | 122 |
| Qy | 185 | GCCATGTTGTGGTGAACCTGAGCAAGTGGCCCTCTGGAGTGGAGGAAGGGCCCCGAG | 244 |
| Db | 121 | GCAATCTTAAGATGTGAGCTTGAGCAAGCTGCTCCTGTGGAGTGGAGGAAGGGTCTCGAG | 62 |
| Qy | 245 | AACCTCAGAGATGGGGACAGATACATCCTCAGGCAGGAGGGGACACAGGTGTGAGCTGCAG | 304 |
| Db | 61 | ACCCTGAAGATGGGGACAGATACACCTCAGGCAGGATGGGGCCGTAATGTGAGCTGCAG | 2 |
| Qy | 305 | A 305 | |
| Db | 1 | A 1 | |

Search completed: September 19, 2004, 22:07:30
Job time : 1127.53 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 19:28:00 ; Search time 199.275 Seconds
(without alignments)
9751.690 Million cell updates/sec

Title: US-10-077-130-4_COPY_10286_10670

Perfect score: 385

Sequence: 1 acatagcctgagcaggagc.....cctcagccagctcaccatc 385

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*

14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 385 | 100.0 | 7564 | 17 | US-10-476-397-15 |
| 2 | 385 | 100.0 | 14061 | 16 | US-10-093-463-73 |
| 3 | 385 | 100.0 | 14109 | 16 | US-10-093-463-71 |
| 4 | 385 | 100.0 | 23907 | 14 | US-10-077-130-6 |
| 5 | 385 | 100.0 | 24120 | 14 | US-10-077-130-4 |
| 6 | 70 | 18.2 | 2534 | 16 | US-10-108-260A-1946 |
| 7 | 46.8 | 12.2 | 81940 | 9 | US-09-759-508B-1 |
| 8 | 46.8 | 12.2 | 81940 | 10 | US-09-960-706-1092 |
| 9 | 46.8 | 12.2 | 81940 | 10 | US-09-873-319-724 |
| 10 | 46.2 | 12.0 | 60 | 10 | US-09-908-975-13516 |
| 11 | 43.4 | 11.3 | 671 | 15 | US-10-184-644-345 |
| 12 | 43.4 | 11.3 | 671 | 15 | US-10-184-644-346 |
| 13 | 43.4 | 11.3 | 1645 | 9 | US-09-726-643-22 |
| 14 | 43.4 | 11.3 | 1645 | 9 | US-10-042-141-22 |

C 15 43.2 11.2 1275 13 US-10-425-114-2127 Sequence 2127, Ap

C 16 43.2 11.2 1547 16 US-10-374-780A-1700 Sequence 1700, Ap

17 42.8 11.1 2254 9 US-09-726-643-44 Sequence 44, Appl

18 42.8 11.1 2254 14 US-10-042-141-44 Sequence 44, Appl

19 41.8 10.9 393 17 US-10-437-963-2910 Sequence 2910, Ap

20 41.8 10.9 716 10 US-09-822-846-117 Sequence 117, App

21 41.8 10.9 2170 16 US-10-264-049-850 Sequence 850, App

22 40.6 10.5 3761 10 US-09-949-029-57 Sequence 57, Appl

C 23 40.6 10.5 7066 10 US-09-919-039-362 Sequence 362, App

C 24 40.2 10.4 951 17 US-10-437-963-63083 Sequence 63083, A

C 25 40.2 10.4 2257 17 US-10-437-963-32032 Sequence 32032, A

26 40 10.4 2433 17 US-10-437-963-92984 Sequence 92984, A

C 27 39.6 10.3 521 17 US-10-437-963-99402 Sequence 99402, A

28 39 10.1 2418 16 US-10-369-493-33579 Sequence 33579, A

29 38.6 10.0 5452 15 US-10-017-161-1481 Sequence 1481, Ap

30 38.6 10.0 5452 16 US-10-292-798-1189 Sequence 1189, Ap

31 38.2 9.9 11358 13 US-09-942-025-16 Sequence 16, Appl

32 38.2 9.9 67311 13 US-09-942-025-16 Sequence 1, Appl

33 37.6 9.8 550 15 US-10-184-644-46 Sequence 46, Appl

34 37.6 9.8 550 15 US-10-184-644-46 Sequence 46, Appl

C 35 37.6 9.8 672 9 US-09-867-550-1397 Sequence 1397, Ap

C 36 37.6 9.8 707 15 US-10-037-270-709 Sequence 709, App

C 37 37.6 9.8 707 16 US-10-117-722-709 Sequence 709, App

38 37.6 9.8 1064 17 US-10-437-963-36517 Sequence 36517, A

39 37.6 9.8 1626 17 US-10-437-963-21289 Sequence 21289, A

40 37.6 9.8 3030 17 US-10-332-947-19 Sequence 19, Appl

41 37.6 9.8 3123 13 US-10-188-186-25 Sequence 25, Appl

C 42 37.6 9.8 3194 16 US-10-108-260A-1776 Sequence 1776, Ap

C 43 37.6 9.8 3457 13 US-10-276-774-1113 Sequence 1113, Ap

C 44 37.6 9.8 4117 13 US-10-433-802-29 Sequence 29, Appl

45 37.6 9.8 6189 13 US-10-342-887-42 Sequence 42, Appl

ALIGNMENTS

RESULT 1

US-10-476-397-15

; Sequence 15, Application US/10476397

; Publication No. US20040115687A1

; GENERAL INFORMATION:

; APPLICANT: YUE, Henry

; APPLICANT: LEE, Ernestine A.

; APPLICANT: DUGGAN, Brendan M.

; APPLICANT: THANGAVELU, Kavitha

; APPLICANT: HONCHELU, Cynthia D.

; APPLICANT: DING, Li

; APPLICANT: JACKSON, Jennifer L.

; APPLICANT: BAUGHN, Mariah R.

; APPLICANT: KALICK, Deborah A.

; APPLICANT: LEE, Sally

; APPLICANT: WARREN, Bridget A.

; APPLICANT: XU, Yuming

; APPLICANT: TRAN, Oyen K.

; APPLICANT: LAL, Preeti G.

; APPLICANT: THORNTON, Michael B.

; APPLICANT: HAFALIA, April J.A.

; APPLICANT: YAO, Monique G.

; APPLICANT: NGUYEN, Damien B.

; APPLICANT: GANDHI, Ameena R.

; APPLICANT: KHAN, Farrah A.

; APPLICANT: CHAWLA, Narinder K.

; APPLICANT: GRIFFIN, Jennifer A.

; APPLICANT: CHINN, Anna M.

; APPLICANT: ELLIOTT, Vicki S.

; APPLICANT: RAMKOMAR, Jayalaxmi

; APPLICANT: ARVIZU, Chandra S.

; APPLICANT: FORSYTHE, Ian J.

; TITLE OF INVENTION: CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS

; FILE REFERENCE: PF-0968 USN

; CURRENT APPLICATION NUMBER: US/10476,397

; CURRENT FILING DATE: 2003-10-31

; PRIOR APPLICATION NUMBER: PCT/US02/13874

;; PRIOR FILING DATE: 2002-05-01
;; PRIOR APPLICATION NUMBER: US 60/288,290
;; PRIOR FILING DATE: 2001-05-02
;; PRIOR APPLICATION NUMBER: US 60/292,468
;; PRIOR FILING DATE: 2001-05-21
;; PRIOR APPLICATION NUMBER: US 60/298,616
;; PRIOR FILING DATE: 2001-06-15
;; PRIOR APPLICATION NUMBER: US 60/301,672
;; PRIOR FILING DATE: 2001-06-28
;; PRIOR APPLICATION NUMBER: US 60/345,008
;; PRIOR FILING DATE: 2002-01-04
;; NUMBER OF SEQ ID NOS: 22
;; SOFTWARE: PERL Program
;; SEQ ID NO 15
;; LENGTH: 7564
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; OTHER INFORMATION: Incyte ID No: 7326129CB1
US-10-476-397-15

Query Match 100.0%; Score 385; DB 17; Length 7564;
Best Local Similarity 100.0%; Pred. No. 7e-98;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACATAGCCTGAGCAGACGGGCTGTGTCGAGCTGCAGATCTGTGGCTGTGGCTGTGGC 60
Db |||||
Qy 4332 ACATAGCCTGAGGAGACGGGCTGTGTCGAGCTGCAGATCTGTGGCTGTGGCTGTGGC 4391
Db |||||

Qy 61 AGATGCTGGGAGTACTCTCTGTGTGTGGGAGGAGACCTCTGCCACTCTCACCGT 120
Db |||||

Qy 4392 AGATGCTGGGAGTACTCTCTGTGTGTGGGAGGAGACCTCTGCCACTCTCACCGT 4451
Db |||||

Qy 121 GAAGCCCTCCAGCCAGTTCACAGAGGCTGTGAGGATGAAGAGCGCTGGAAGGGC 180
Db |||||

Qy 4452 GAAGCCCTCCAGCCAGTTCACAGAGGCTGTGAGGATGAAGAGCGCTGGAAGGGC 4511
Db |||||

Qy 181 CACAGCCATGTTGTGTGTAATCTGAGCAAGTGGCCCTGTGCACTGGAGGAAGGGCC 240
Db |||||

Qy 4512 CACAGCCATGTTGTGTGTAATCTGAGCAAGTGGCCCTGTGCACTGGAGGAAGGGCC 4571
Db |||||

Qy 241 CGAGAACTCAGAGATGGGACAGATACATCTTGAAGGAGGAGGACAGGTGTGAGCT 300
Db |||||

Qy 4572 CGAGAACTCAGAGATGGGACAGATACATCTTGAAGGAGGAGGAGGAGGAGGAGG 4631
Db |||||

Qy 301 GCAGATCTGTGGCTGSCCATGCGGACCGCGGGGAGTACTTGTGTGTGGGGCAGGA 360
Db |||||

Qy 4632 GCAGATCTGTGGCTGSCCATGCGGACCGCGGGGAGTACTTGTGTGTGGGGCAGGA 4691
Db |||||

Qy 361 GAGGACCTCAGCCAGCTCACCATC 385
Db |||||

Qy 4692 GAGGACCTCAGCCAGCTCACCATC 4716
Db |||||

RESULT 2

US-10-093-463-73
;; Sequence 73, Application US/10093463
;; Publication NO. US20030208039A1
;; GENERAL INFORMATION:
;; APPLICANT: Padigar, Muralidhara
;; APPLICANT: Shenoy, Suresh
;; APPLICANT: Kekuda, Ramesh
;; APPLICANT: Gusev, Vladimir
;; APPLICANT: Pochart, Pascal
;; APPLICANT: Zhong, Mei
;; APPLICANT: Rastelli, Luca
;; APPLICANT: Mezes, Peter
;; APPLICANT: Smithson, Glenda
;; APPLICANT: Guo, Xiaojia
;; APPLICANT: Gerlach, Valerie
;; APPLICANT: Casman, Stacie
;; APPLICANT: Boldog, Ferenc

;; APPLICANT: Li, Li
;; APPLICANT: Zerhusen, Bryan
;; APPLICANT: Tchernev, Velizar
;; APPLICANT: Gangolli, Esha
;; APPLICANT: Vernet, Corine
;; APPLICANT: Pena, Carol
;; APPLICANT: Burgess, Catherine
;; APPLICANT: Liu, Xiaohong
;; APPLICANT: Spytek, Kimberly
;; APPLICANT: Gorman, Linda
;; APPLICANT: Spaderna, Steven
;; APPLICANT: Voss, Edward
;; APPLICANT: Malyankar, Uriel
;; APPLICANT: Anderson, David
;; APPLICANT: Patturajan, Meera
;; APPLICANT: Miller, Charles
;; APPLICANT: Taupier, Raymond J. Jr.
;; TITLE OF INVENTION: No. US20030208039A1 Antibodies that Bind to Antigenic Polypepti
;; FILE REFERENCE: 21402-290A (Cura 590AT)
;; CURRENT APPLICATION NUMBER: US/10/093,463
;; CURRENT FILING DATE: 2002-06-24
;; PRIOR APPLICATION NUMBER: 60/283,675
;; PRIOR FILING DATE: 2001-04-14
;; PRIOR APPLICATION NUMBER: 60/338,092
;; PRIOR FILING DATE: 2001-12-03
;; PRIOR APPLICATION NUMBER: 60/274,281
;; PRIOR FILING DATE: 2001-03-08
;; PRIOR APPLICATION NUMBER: 60/274,101
;; PRIOR FILING DATE: 2001-03-08
;; PRIOR APPLICATION NUMBER: 60/325,681
;; PRIOR FILING DATE: 2001-09-27
;; PRIOR APPLICATION NUMBER: 60/304,354
;; PRIOR FILING DATE: 2001-07-10
;; PRIOR APPLICATION NUMBER: 60/279,995
;; PRIOR FILING DATE: 2001-03-30
;; PRIOR APPLICATION NUMBER: 60/294,999
;; PRIOR FILING DATE: 2001-05-31
;; PRIOR APPLICATION NUMBER: 60/287,424
;; PRIOR FILING DATE: 2001-04-30
;; PRIOR APPLICATION NUMBER: 60/299,027
;; PRIOR FILING DATE: 2001-06-18
;; PRIOR APPLICATION NUMBER: 60/309,198
;; PRIOR FILING DATE: 2001-07-31
;; PRIOR APPLICATION NUMBER: 60/281,194
;; PRIOR FILING DATE: 2001-04-04
;; PRIOR APPLICATION NUMBER: 60/274,194
;; PRIOR FILING DATE: 2001-03-08
;; PRIOR APPLICATION NUMBER: 60/274,849
;; PRIOR FILING DATE: 2001-03-09
;; PRIOR APPLICATION NUMBER: 60/330,380
;; PRIOR FILING DATE: 2001-10-18
;; PRIOR APPLICATION NUMBER: 60/275,235
;; PRIOR FILING DATE: 2001-03-12
;; PRIOR APPLICATION NUMBER: 60/288,342
;; PRIOR FILING DATE: 2001-05-03
;; PRIOR APPLICATION NUMBER: 60/275,578
;; PRIOR FILING DATE: 2001-03-13
;; NUMBER OF SEQ ID NOS: 370
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 73
;; LENGTH: 14061
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (15) .. (14039)
US-10-093-463-73

Query Match 100.0%; Score 385; DB 16; Length 14061;
Best Local Similarity 100.0%; Pred. No. 7e-98;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACATAGCCTGAGGACGAGCGGGCTGTGTGCGAGCTGCAGATCTGTGGCCTGGCTGTGGC 60
Db 2447 ACATAGCCTGAGGACGAGCGGGCTGTGTGCGAGCTGCAGATCTGTGGCCTGGCTGTGGC 2506
Qy 61 AGATGCTGGGAGTACTCTGTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCGT 120
Db 2507 AGATGCTGGGAGTACTCTGTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCGT 2566
Qy 121 GAAGGCCCTGCACGCCAAGTTTCACAGAGGGCTCTGAGGAATGAAGAGCGCCGTGGAAGGGGC 180
Db 2567 GAAGGCCCTGCACGCCAAGTTTCACAGAGGGCTCTGAGGAATGAAGAGCGCCGTGGAAGGGGC 2626
Qy 181 CACAGCCATGTTGTGTGTGTGAACTGAGCAAGGTGGCCCTGTGGAGTGGAGGAGGGGCC 240
Db 2627 CACAGCCATGTTGTGTGTGTGAACTGAGCAAGGTGGCCCTGTGGAGTGGAGGAGGGGCC 2686
Qy 241 CGAGAACCTCAGAGATGGGGACAGATACATCTTGAGCAGAGGGCTGAGGAGGAGGAGTGTAGCT 300
Db 2687 CGAGAACCTCAGAGATGGGGACAGATACATCTTGAGCAGAGGGCTGAGGAGGAGGAGTGTAGCT 2746
Qy 301 GCAGATCTGTGGCTGGCCATGGCGGAGCGCCGGGAGTACTTGTGTGTGTGGCGGACAGGA 360
Db 2747 GCAGATCTGTGGCTGGCCATGGCGGAGCGCCGGGAGTACTTGTGTGTGTGGCGGACAGGA 2806
Qy 361 GAGGACCTCAGCCACGCTCACCATC 385
Db 2807 GAGGACCTCAGCCACGCTCACCATC 2831

RESULT 3

US-10-093-463-71
; Sequence 71, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esba
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypepti
; FILE REFERENCE: 21402-290A (Cura 550Aa)
; CURRENT APPLICATION NUMBER: US/10/093,463
; FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03

; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 71
; LENGTH: 14109
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15)..(14088)
US-10-093-463-71
Query Match 100.0%; Score 385; DB 16; Length 14109;
Best Local Similarity 100.0%; Pred. No. 7e-98;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACATAGCCTGAGGACGAGCGGGCTGTGTGCGAGCTGCAGATCTGTGGCCTGGCTGTGGC 60
Db 2447 ACATAGCCTGAGGACGAGCGGGCTGTGTGCGAGCTGCAGATCTGTGGCCTGGCTGTGGC 2506
Qy 61 AGATGCTGGGAGTACTCTGTGTGTGTGGGAGGAGGAGGACCTCTGCCACTCTCACCGT 120
Db 2507 AGATGCTGGGAGTACTCTGTGTGTGTGGGAGGAGGAGGACCTCTGCCACTCTCACCGT 2566
Qy 121 GAAGGCCCTGCACGCCAAGTTTCACAGAGGGCTCTGAGGAATGAAGAGCGCCGTGGAAGGGGC 180
Db 2567 GAAGGCCCTGCACGCCAAGTTTCACAGAGGGCTCTGAGGAATGAAGAGCGCCGTGGAAGGGGC 2626
Qy 181 CACAGCCATGTTGTGTGTGTGAACTGAGCAAGGTGGCCCTGTGGAGTGGAGGAGGGGCC 240
Db 2627 CACAGCCATGTTGTGTGTGTGAACTGAGCAAGGTGGCCCTGTGGAGTGGAGGAGGGGCC 2686
Qy 241 CGAGAACCTCAGAGATGGGGACAGATACATCTTGAGCAGAGGGCTGAGGAGGAGGAGTGTAGCT 300
Db 2687 CGAGAACCTCAGAGATGGGGACAGATACATCTTGAGCAGAGGGCTGAGGAGGAGGAGTGTAGCT 2746
Qy 301 GCAGATCTGTGGCTGGCCATGGCGGAGCGCCGGGAGTACTTGTGTGTGTGGCGGACAGGA 360
Db 2747 GCAGATCTGTGGCTGGCCATGGCGGAGCGCCGGGAGTACTTGTGTGTGTGGCGGACAGGA 2806
Qy 361 GAGGACCTCAGCCACGCTCACCATC 385


```
Db 2807 GAGGACCTCAGCCACGCTCACCATC 2831

RESULT 4
US-10-077-130-6
; Sequence 6, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1RCP1(N)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6

Query Match 100.0%; Score 385; DB 14; Length 23907;
Best Local Similarity 100.0%; Pred. No. 7e-98;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACATAGCTGAGGACGAGGCGGCTGTGCGAGCTGCAGATCTGTGGCTGGCTGTGGC 60
Db 10215 ACATAGCTGAGGACGAGGCGGCTGTGCGAGCTGCAGATCTGTGGCTGGCTGTGGC 10274
QY 61 AGATGCTGGGAGPACTCTCTGTGTGTGTGGGAGGAGGACCTTGCCACTCTCACCGT 120
Db 10275 AGATGCTGGGAGPACTCTCTGTGTGTGTGGGAGGAGGACCTTGCCACTCTCACCGT 10334
QY 121 GAGGCGCTGCCAGCCCAAGTTCACAGAGGCTCTGAGGAATGAGAGGCGTGAAGGGGC 180
Db 10335 GAGGCGCTGCCAGCCCAAGTTCACAGAGGCTCTGAGGAATGAGAGGCGTGAAGGGGC 10394
QY 181 CACAGCCATGTTGTGTGTGAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGGC 240
Db 10395 CACAGCCATGTTGTGTGTGAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGGC 10454
QY 241 CGAGAACTCAGAGATGGGACAGATACATCTTGAGCGAGGAGGACCGTGTGAGCT 300
Db 10455 CGAGAACTCAGAGATGGGACAGATACATCTTGAGCGAGGAGGACCGTGTGAGCT 10514
QY 301 GCAGATCTGTGGCTGGCCATGGCGACGCGCGGGAGTACTTGTGTGTGTGGCGGAGGA 360
Db 10515 GCAGATCTGTGGCTGGCCATGGCGACGCGCGGGAGTACTTGTGTGTGTGGCGGAGGA 10574
QY 361 GAGGACCTCAGCCACGCTCACCATC 385
Db 10575 GAGGACCTCAGCCACGCTCACCATC 10599

RESULT 5
US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1RCP1(N)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-4

Query Match 100.0%; Score 385; DB 14; Length 24120;
Best Local Similarity 100.0%; Pred. No. 7e-98;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACATAGCTGAGGACGAGGCGGCTGTGCGAGCTGCAGATCTGTGGCTGGCTGTGGC 60
Db 10286 ACATAGCTGAGGACGAGGCGGCTGTGCGAGCTGCAGATCTGTGGCTGGCTGTGGC 10345
QY 61 AGATGCTGGGAGTACTCTCTGTGTGTGTGGGAGGAGGACCTTGCCACTCTCACCGT 120
Db 10346 AGATGCTGGGAGTACTCTCTGTGTGTGTGGGAGGAGGACCTTGCCACTCTCACCGT 10405
QY 121 GAGGCGCTGCCAGCCCAAGTTCACAGAGGCTCTGAGGAATGAGAGGCGTGAAGGGGC 180
Db 10406 GAGGCGCTGCCAGCCCAAGTTCACAGAGGCTCTGAGGAATGAGAGGCGTGAAGGGGC 10465
QY 181 CACAGCCATGTTGTGTGTGAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGGC 240
Db 10466 CACAGCCATGTTGTGTGTGAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGGC 10525
QY 241 CGAGAACTCAGAGATGGGACAGATACATCTTGAGCGAGGAGGACCGTGTGAGCT 300
Db 10526 CGAGAACTCAGAGATGGGACAGATACATCTTGAGCGAGGAGGACCGTGTGAGCT 10585
QY 301 GCAGATCTGTGGCTGGCCATGGCGACGCGCGGGAGTACTTGTGTGTGTGGCGGAGGA 360
Db 10586 GCAGATCTGTGGCTGGCCATGGCGACGCGCGGGAGTACTTGTGTGTGTGGCGGAGGA 10645
QY 361 GAGGACCTCAGCCACGCTCACCATC 385
Db 10646 GAGGACCTCAGCCACGCTCACCATC 10670

RESULT 6
US-10-108-260A-1946
; Sequence 1946, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1946
; LENGTH: 2534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1946

Query Match 18.2%; Score 70; DB 16; Length 2534;
Best Local Similarity 51.4%; Pred. No. 5.7e-10;
Matches 188; Conservative 0; Mismatches 175; Indels 3; Gaps 1;

QY 17 GACGGGGCTGTGTGCGAGCTGCAGATCTGTGGCTTGCTGTGGCAGATGCTGGGAGTAC 76
Db 937 GAGGGCACGATGGCCATGCTGGTTCATCCGGGGGCTTCGCTCAAGACGCGGCGAGTAC 996
```

QY 77 TCCTGTGTGTGTGGGAGGAGGACCTCTGCACCTCTCACCCTGAAGCCCTGCCAGCC 136
 Db 997 ACGTGTAGTGTGAGGCTTCCAGAGACAGCCAGCTTCAATGTGAAGAAAGCAAC 1056
 QY 137 AAGTTACAGAGGCTGTGAAGATGAAGAGCCGCTGGAAGGGCCACAGCCATGTTGTGG 196
 Db 1057 TGCTTACAGAGGCTGACCAATCTGCAGGTGGAGGAAAGGCACAGCTGTGTTCAAG 1116
 QY 197 TGTGA---ACTGAGCAAGTGGCCCTGTGTGAGTGGAGGAAGGGCCCGAGAACCTCAGA 253
 Db 1117 TGCAAGACGAGACCCCGGGCCACAGTACCTGGCGCAAGGCTCTTGGAGCTACGG 1176
 QY 254 GATGGGACAGATACATCTCTGAGCGAGGGGACCAAGGTGTGAGCTGCAGATCTCTGGC 313
 Db 1177 GCCTCAGGAAGACACAGCCAGCCAGGAGGGCTGACCCCTGCGGTCCACCATCAGTGGC 1236
 QY 314 CTGGCCATGGGAGCGCCGGGAGTACTTGTGTGTGTGGCGGAGAGGAGGAGGAGGAGG 373
 Db 1237 CTGGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1296
 QY 374 ACGCTC 379
 Db 1297 CAGCTC 1302

RESULT 7
 US-09-759-508B-1
 ; Sequence 1, Application US/09759508B
 ; Publication No. US20020182599A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fishman, Mark C.
 ; TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease
 ; FILE REFERENCE: 00786/381002
 ; CURRENT APPLICATION NUMBER: US/09/759,508B
 ; CURRENT FILING DATE: 2001-01-12
 ; PRIOR APPLICATION NUMBER: US 60/175,787
 ; PRIOR FILING DATE: 2000-01-12
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 81940
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (133)..(80910)
 ; OTHER INFORMATION:
 US-09-759-508B-1

Query Match 12.2%; Score 46.8; DB 9; Length 81940;
 Best Local Similarity 48.8%; Pred. No. 0.0017;
 Matches 157; Conservative 0; Mismatches 162; Indels 3; Gaps 1;
 QY 66 CTGGGAGTACTCTCTGTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 125
 Db 16091 CAGGAGAGGTTTCTCTCAGGCTGCTAATGCCAAATCTGCAGGCAATCTGAAAGTGAAG 16150
 QY 126 CCTGCCAGCCAGTTTCAGAGGCTCTGAGGAAATGAAGAGCCGCTGGAAGGGCCACAG 185
 Db 16151 AATTGCCTCTTATCTCATCACCTCTCAGTGTGTTAAAGTCTTCGAGAAAGATGAGG 16210
 QY 186 CCATGTTGTGTGTAACCTGAGCAAGGTGGCC---CCTGTGGAGTGGAGGAAGGGGCCG 242
 Db 16211 CTAAGTTTGTGAGTGTGAAGTATCCAGGAGGCCAAACATTCCGTTGGCTTAAAGGAACCC 16270
 QY 243 AGAACCTCAGAGATGGGACAGATACCTCTGAGCGAGGAGGAGGAGGAGGAGGAGGAGG 302
 Db 16271 AGGAATCAGAGTGTGAGCTTATAAAGATGGCAGCTAAGCAATTAAGTGG 16330
 QY 303 AGATCTGTGGCTGGCCATGGCGGACGCGGGGAGTACTTGTGTGTGCGGGCAGGAGA 362
 Db 16331 TGATCAAGTCAGCTGCTTTTGAAGATGAAGCAAAATACATGTTTGAAGCTGAAGATAAGC 16390

QY 363 GGACCTCAGCCACGCTCACCAT 384
 Db 16391 ACACAAGTGGCAAACTGATCAT 16412

RESULT 8
 US-09-960-706-1092
 ; Sequence 1092, Application US/09960706
 ; Publication No. US20030134280A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Munger, William E.
 ; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperpla
 ; FILE REFERENCE: 44921-5029-01US
 ; CURRENT APPLICATION NUMBER: US/09/960,706
 ; CURRENT FILING DATE: 2001-09-24
 ; PRIOR APPLICATION NUMBER: 60/223,323
 ; PRIOR FILING DATE: 2000-08-07
 ; PRIOR APPLICATION NUMBER: 09/873,319
 ; PRIOR FILING DATE: 2001-06-05
 ; NUMBER OF SEQ ID NOS: 1124
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 1092
 ; LENGTH: 81940
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20030134280A1 X90568
 US-09-960-706-1092

Query Match 12.2%; Score 46.8; DB 10; Length 81940;
 Best Local Similarity 48.8%; Pred. No. 0.0017;
 Matches 157; Conservative 0; Mismatches 162; Indels 3; Gaps 1;
 QY 66 CTGGGAGTACTCTCTGTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 125
 Db 16091 CAGGAGAGGTTTCTCTCAGGCTGCTAATGCCAAATCTGCAGGCAATCTGAAAGTGAAG 16150
 QY 126 CCTGCCAGCCAGTTTCAGAGGCTCTGAGGAAATGAAGAGCCGCTGGAAGGGCCACAG 185
 Db 16151 AATTGCCTCTTATCTCATCACCTCTCAGTGTGTTAAAGTCTTCGAGAAAGATGAGG 16210
 QY 186 CCATGTTGTGTGTAACCTGAGCAAGGTGGCC---CCTGTGGAGTGGAGGAAGGGGCCG 242
 Db 16211 CTAAGTTTGTGAGTGTGAAGTATCCAGGAGGCCAAACATTCCGTTGGCTTAAAGGAACCC 16270
 QY 243 AGAACCTCAGAGATGGGACAGATACCTCTGAGCGAGGAGGAGGAGGAGGAGGAGGAGG 302
 Db 16271 AGGAATCAGAGTGTGAGCTTATAAAGATGGCAGCTAAGCAATTAAGTGG 16330
 QY 303 AGATCTGTGGCTGGCCATGGCGGACGCGGGGAGTACTTGTGTGTGCGGGCAGGAGA 362
 Db 16331 TGATCAAGTCAGCTGCTTTTGAAGATGAAGCAAAATACATGTTTGAAGCTGAAGATAAGC 16390
 QY 363 GGACCTCAGCCACGCTCACCAT 384
 Db 16391 ACACAAGTGGCAAACTGATCAT 16412

RESULT 9
 US-09-873-319-724
 ; Sequence 724, Application US/09873319A
 ; Publication No. US20030134324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Munger, William E.
 ; APPLICANT: Kulkarni, Prakash
 ; APPLICANT: Getzenberg, Robert H.
 ; APPLICANT: Waga, Iwao
 ; APPLICANT: Yamamoto, Jun
 ; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
 ; FILE REFERENCE: 44921-5029-US
 ; CURRENT APPLICATION NUMBER: US/09/873,319A

```
; CURRENT FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; EARLIER FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 724
; LENGTH: 81940
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 X90568
US-09-873-319-724

Query Match      12.2%; Score 46.8; DB 10; Length 81940;
Best Local Similarity 48.8%; Pred. No. 0.0017;
Matches 157; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

QY 66 CTGGGGAGTACTCTGTGTGTGGGAGGAGGAGGACTCTGCCACTCTCACCGTGAAGG 125
DB 16091 CAGGAGAGTTCTCTCCAGGCTCTAATGCCAATCTGCAGCCATCTGAAAGTGAAG 16150

QY 126 CCCTGCCAGCCAACTTCCAGAGGCTCTGAGGAATGAAGAGCGGTGGAAGGGGCCACAG 185
DB 16151 AATTGCGCTTATCTTCATCACACCTCTCAGTGATGTTAAAGTCTTCGAGAAAGATGAGG 16210

QY 186 CCATGTTGTGTGTAAGTACAGCAAGTGGCC---CCTGTGGAGTGGAGGAGGCGCCG 242
DB 16211 CTAAGTTGAGTGTGAAGTATCCAGGAGGCCAAACATTCCTGTGGTAAAGGAACCC 16270

QY 243 AGAACCTCAGAGATGGGAGCAGATACATCTCTGAGGAGGAGGAGGAGGAGGAGTGTGAGCTGC 302
DB 16271 AGAAATCACAGGTGATGACAGATTGAGCTTATAAGGATGCACTTAAGCATTCATG 16330

QY 303 AGATCTGTGGCTGGCCATGGGAGCCGGGAGTACTGTGTGTGTGGCGGCAGGAGA 362
DB 16331 TGATCAAGTACGTGCTTTTGAAGATGAAGCAAAATACATGTTGAAGCTGAAGATAAGC 16390

QY 363 GGACCTCAGCCACGCTCACCAT 384
DB 16391 ACACAAGTGGCAACTGATCAT 16412

RESULT 10
US-09-908-975-13516
; Sequence 13516, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICING
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13516
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-13516

Query Match      12.0%; Score 46.2; DB 10; Length 60;
Best Local Similarity 86.4%; Pred. No. 0.0025;
Matches 51; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
QY 125 GGCCTGCCAGCAAGTTTCACAGAGGCTCTGAGGAATGAAGAGCGCTGGAAGGGGCCAC 183
DB 1 GGCATGCCTTCCAAGTTTCATAGAGGCTCTGAGGAATGAAGAGCGCCACAGAGGGGACAC 59

RESULT 11
US-10-184-644-346
; Sequence 346, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P34301C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 346
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-184-644-346

Query Match      11.3%; Score 43.4; DB 15; Length 671;
Best Local Similarity 7.2%; Pred. No. 0.015;
Matches 18; Conservative 110; Mismatches 122; Indels 0; Gaps 0;

QY 10 GAGGAGGAGCGGGCTGTGTGCGAGTGCAGATCTGTGGCTGTGCTGTGCAGATCTGTG 69
DB 163 KMSVKRARKASDLDQASVSPSEENSESSSEKTSDDQFTPEKKAAPRPRGLGG 222

QY 70 GGAGTACTCTCTGTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 129
DB 223 RKXKAPASDSDSKADSDGAKPEPVAVARSASSSSSSSSSSSDSDSVKPKPRGRPAEK 282

QY 130 GCCAGCCAAAGTTCACAGAGGCTGTGAGGAATGAAGAGCGCTGTGGAAGGGGCCACAGCCAT 189
DB 283 PLPKRGRKPKPRPPSSSSSSSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 342

QY 190 GTTGTGTGTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 249
DB 343 REQEKEKERRRRERADRGAEERGGGSGGDELREDDDEPVRKGRKGRGRGPPSSSSSEPE 402

QY 250 CAGAGATGGG 259
DB 403 AELEREAKKS 412

RESULT 12
US-10-184-634-346
; Sequence 346, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
```

```
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 346
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-346

Query Match      11.3%; Score 43.4; DB 15; Length 671;
Best Local Similarity 7.2%; Pred. No. 0.015;
Matches 18; Conservative 110; Mismatches 122; Indels 0; Gaps 0;

QY 10 GAGCAGGACGGGCTGTGTCGAGCTGCAGATCTGTGGCTGTGGCTGTGGCAGATCTGTG 69
Db 163 KMSVSKRKAKASDLDQASVSPSEENSESESESEKTSDDFTPEKKAARVPRGFLGG 222
QY 70 GGAGTACTCTCTGTGTGTGGGAGGAGAGACCTCTGCCACTCTCACCCTGAAGGCCCT 129
Db 223 RKKKKAPSADSDSKADSDGAKPEPVAMARSASSSSSSSDSDSVKPPRKRPAEX 282
QY 130 GCCAGCCAGTTCACAGAGGTCTGAGAGATGAAGAGCCGTGTGAAGGGGCCACAGCCAT 189
Db 283 PLPGRKPKRPPPPSSSSSDSDSDVDRISWKRRDEARRRLEARRRQEEELRL 342
QY 190 GTTGTGTGAAGTCACTGAGCAGGTGGCCCTGTGAGTGGAGGAGGGGCCGAGAACCT 249
Db 343 REQEKERERRRADRGEERGGSGGDELREDDEPVKXKRKRGRRGPPSSDSEPE 402
QY 250 CAGAGATGGG 259
Db 403 AELEREAKKS 412
```

```
RESULT 13
US-09-726-643-22.
; Sequence 22, Application US/09726643
; Patent No. US2002028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: PZ040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR FILING DATE: 2000-12-01
; PRIOR FILING DATE: 2000-06-02
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-726-643-22

Query Match      11.3%; Score 43.4; DB 9; Length 1645;
Best Local Similarity 48.3%; Pred. No. 0.015;
Matches 187; Conservative 0; Mismatches 191; Indels 9; Gaps 2;
```

```
QY 8 CTGAGGAGGAGGCGGCTGTGTGGAGCTGCAGATCTGTGGCTGTGGCTGTGGCAGATGCT 67
Db 950 CTGAGAGATGAGGGGCCCCATCGCGGCTGTGTGGCTGTGGCTGTGGCAGGCGG 1009
QY 68 GGGGAGTACTCTCTGTGTGTGGGAGGAGGAGACCTCTGCCACTCTCACCCTGAAGGCC 127
Db 1010 GGCAGATTTCAGTGGCTGTGGAGATGAGTGTGCTTCACTGTCACTATCACCATCAGAC 1069
```

```
QY 128 CTGCCAGCCAGTTACAGAGGGTCTGAGGAATGAAGAGCGCGTGAAG-GGGCCACAGC 186
Db 1070 GTCTCTCTCTGTGATCGTGTATCCAGCGCAGAGGTGTATGTGGCAGCGCTGCGCTGGAG 1129
QY 187 CATGTTGTGTGTGAACCTGAGCAAGGTGCGCCCTGTGGAG-----TGGAGGAAGGGG 238
Db 1130 CGTGTGGTGTGACCTGTGAGCTATATCGCGGCTGTGGCAGAGGTGCGTGTGGACCAAGGAT 1189
QY 239 CCGGAGAACCTCAGAGATGGGACAGATACATCTCTGAGGACGAGGGGACCAAGGTGTGAG 298
Db 1190 GGAGAGGAGGTGTGTGGAGAGCGCGCTGCTCTTCAGAGAGAGACACTGTTCGCCCGC 1249
QY 299 CTGAGATCTGTGGCTGTGGCCATGCGGAGCGCGGGAGTACTTGTGTGTGTGGGGCAG 358
Db 1250 CTGGTCTGCGCGCTGTCCAGCTCGAGGACTCGGCGAGTACTTGTGTGAAATTGACGAT 1309
QY 359 GAGAGGACCTCAGCCAGCGCTCACCATC 385
Db 1310 GAGTGGCGCTCTTCACTGTCAACGTC 1336
```

```
RESULT 14
US-10-042-141-22
; Sequence 22, Application US/10042141
; Publication No. US20020183503A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: PZ040P1
; CURRENT APPLICATION NUMBER: US/10/042,141
; PRIOR FILING DATE: 2002-01-11
; PRIOR FILING DATE: 1999-06-07
; PRIOR FILING DATE: 2000-12-01
; PRIOR FILING DATE: 2000-06-02
; PRIOR FILING DATE: 2000-06-02
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1645
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-141-22
```

```
Query Match      11.3%; Score 43.4; DB 14; Length 1645;
Best Local Similarity 48.3%; Pred. No. 0.015;
Matches 187; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

QY 8 CTGAGGAGGAGGCGGCTGTGTGGAGCTGCAGATCTGTGGCTGTGGCTGTGGCAGATGCT 67
Db 950 CTGAGAGATGAGGGGCCCCATCGCGGCTGTGTGGCTGTGGCTGTGGCAGGCGG 1009
QY 68 GGGGAGTACTCTCTGTGTGTGGGAGGAGGAGACCTCTGCCACTCTCACCCTGAAGGCC 127
Db 1010 GGCAGATTTCAGTGGCTGTGGAGATGAGTGTGCTTCACTGTCACTATCACCATCAGAC 1069
QY 128 CTGCCAGCCAGTTACAGAGGGTCTGAGGAATGAAGAGCGCGTGGAG-GGGCCACAGC 186
Db 1070 GTCTCTCTGTGATCTGTATCCAGCGCAGAGGTGTATGTGGCAGCGCTGCGCTGGAG 1129
QY 187 CATGTTGTGTGTGAACCTGAGCAAGGTGCGCCCTGTGGAG-----TGGAGGAAGGGG 238
Db 1130 CGTGTGGTGTGACCTGTGAGCTATGCGCGGCTGTGGCAGAGGTGCGTGTGGACCAAGGAT 1189
QY 239 CCGGAGAACCTCAGAGATGGGACAGATACATCTTCAGGAGGAGGGGACCAAGGTGTGAG 298
Db 1190 GGAGAGGAGGTGTGTGGAGAGCGCGCTGTCTTCTSCAGAGAGAGACACTGTTCGCCCGC 1249
QY 299 CTGAGATCTGTGGCTGTGGCCATGCGGAGCGCGGGAGTACTTGTGTGTGTGGGGCAG 358
Db 1250 CTGGTGTCTGCGCTGTCCAGCTCGAGGACTCCGCGAGTACTTGTGTGAAATTGACGAT 1309
```

Qy 359 GAGGAGCCTCAGCCAGCTCACCATC 385
Db 1310 GAGTCGGCTCTTCACTGTCAACGTC 1336

RESULT 15

US-10-425-114-2127/c
; Sequence 2127, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 2127
; LENGTH: 1275
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700204307_FLI
US-10-425-114-2127

Query Match 11.2%; Score 43.2; DB 13; Length 1275;
Best Local Similarity 48.0%; Pred. No. 0.017;
Matches 123; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
Qy 121 GAAGGCCCTGCCAGCCAAAGTTTCACAGAGAGGTCTGAGGAATGAAGAGGCCGTGGAAGGGC 180
Db 443 GAATGCTCACCGCCATGGCCATGATGGTATGTCGAGGAGAGGCCGTGAGGGGAG 384
Qy 181 CACAGCCATGTTGGTGTGAACCTGACAAAGGTGGCCCTGTGGAGTGGAGGAAGGGCC 240
Db 383 CACATTGGTGTTCATGACGGAGAGGCCGCGCTCCACCTCGGAGAGGGCGCGATGAG 324
Qy 241 CGAGAACTCAGAGATGGGACAGATACATCTTGAGCGAGGGGACAGGTGTGAGCT 300
Db 323 CACCCCTTGGGTTGTCAGTGGACCTTACACAGACCGTGGCGCCGACAGCCGGC 264
Qy 301 GCAGATCTGTGGCTGGCCATGGCGACGCCGGGGAGTACTTGTGTGTGGGGCAGGA 360
Db 263 CTCGATCTCCGGCAGAGCGTGGGAGGCCCTCGGGCCGCCACGAGTTGTCTCGTCAGGA 204
Qy 361 GAGGACCTCAGCCACG 376
Db 203 GGAGCCCTCGTCG 188

Search completed: September 20, 2004, 02:51:57
Job time : 201.275 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:09:24 : Search time 34.5752 Seconds
(without alignments)
6179.453 Million cell updates/sec

Title: US-10-077-130-4_COPY_10286_10670

Perfect score: 385
Sequence: 1 acatagcctgaggcaggacg.....cctcagccagctcaccatc 385

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| C 1 | 54.2 | 14.1 | 7218 | 1 | US-08-232-463-14 |
| C 2 | 44.8 | 11.6 | 289 | 3 | US-09-007-005-17 |
| C 3 | 44.8 | 11.6 | 289 | 3 | US-09-244-796-17 |
| C 4 | 40.6 | 10.5 | 7055 | 4 | US-09-976-594-941 |
| C 5 | 39 | 10.1 | 2277 | 1 | US-08-676-967-5 |
| C 6 | 39 | 10.1 | 2277 | 1 | US-08-676-974-5 |
| C 7 | 39 | 10.1 | 2277 | 2 | US-09-098-487-5 |
| C 8 | 37.6 | 9.8 | 707 | 4 | US-09-620-312D-709 |
| C 9 | 37.4 | 9.7 | 1975 | 4 | US-09-308-345A-6 |
| C 10 | 37.2 | 9.7 | 9531 | 4 | US-09-163-748C-3 |
| C 11 | 37.2 | 9.7 | 11461 | 3 | US-08-669-161A-29 |
| C 12 | 37 | 9.6 | 2478 | 4 | US-09-687-050-5 |
| C 13 | 37 | 9.6 | 2478 | 4 | US-09-921-667-15 |
| C 14 | 37 | 9.6 | 3597 | 4 | US-09-016-434-1319 |
| C 15 | 37 | 9.6 | 3597 | 4 | US-09-023-655-1314 |
| C 16 | 36 | 9.4 | 1572 | 4 | US-09-252-991A-3270 |
| C 17 | 36 | 9.4 | 1965 | 4 | US-09-252-991A-3234 |
| C 18 | 36 | 9.4 | 2325 | 4 | US-09-252-991A-3413 |
| C 19 | 35.6 | 9.2 | 1926 | 4 | US-09-249-585A-4 |
| C 20 | 35.6 | 9.2 | 1931 | 2 | US-09-130-114-2 |
| C 21 | 35.4 | 9.2 | 923 | 4 | US-09-336-643A-15 |
| C 22 | 35.4 | 9.2 | 2612 | 3 | US-09-042-785A-3 |
| C 23 | 35.4 | 9.2 | 2638 | 3 | US-09-042-785A-22 |
| C 24 | 35.4 | 9.2 | 3474 | 4 | US-09-527-236A-1 |
| C 25 | 35.4 | 9.2 | 3474 | 4 | US-09-756-854-1 |
| C 26 | 35 | 9.1 | 584 | 3 | US-08-959-382-3 |
| C 27 | 35 | 9.1 | 584 | 4 | US-09-314-844F-3 |

| | | | | | | |
|------|------|-----|-------|---|---------------------|-------------------|
| 28 | 35 | 9.1 | 33529 | 3 | US-09-144-085-3 | Sequence 3, Appli |
| 29 | 34.6 | 9.0 | 1479 | 1 | US-08-644-271-31 | Sequence 31, Appl |
| 30 | 34.6 | 9.0 | 1479 | 4 | US-09-077-955-35 | Sequence 35, Appl |
| C 31 | 34.2 | 8.9 | 2079 | 4 | US-09-992-481-1 | Sequence 1, Appli |
| C 32 | 33.8 | 8.8 | 3277 | 4 | US-09-963-137-163 | Sequence 163, App |
| C 33 | 33.6 | 8.7 | 5103 | 4 | US-09-252-991A-5192 | Sequence 5192, Ap |
| C 34 | 33.6 | 8.7 | 6876 | 4 | US-09-252-991A-5283 | Sequence 5283, Ap |
| C 35 | 33.4 | 8.7 | 3695 | 3 | US-09-211-704A-1 | Sequence 1, Appli |
| C 36 | 33.2 | 8.6 | 2277 | 1 | US-08-676-967-2 | Sequence 2, Appli |
| C 37 | 33.2 | 8.6 | 2277 | 1 | US-08-676-974-2 | Sequence 2, Appli |
| C 38 | 33.2 | 8.6 | 2277 | 2 | US-09-428-487-2 | Sequence 2, Appli |
| C 39 | 33.2 | 8.6 | 50937 | 3 | US-09-428-517-1 | Sequence 1, Appli |
| C 40 | 33 | 8.6 | 3282 | 1 | US-08-276-852-154 | Sequence 154, App |
| C 41 | 33 | 8.6 | 3282 | 1 | US-08-276-852-154 | Sequence 169, App |
| C 42 | 33 | 8.6 | 3282 | 1 | US-08-899-575-154 | Sequence 154, App |
| C 43 | 33 | 8.6 | 3282 | 1 | US-08-899-575-154 | Sequence 169, App |
| C 44 | 33 | 8.6 | 3282 | 1 | US-08-899-575-154 | Sequence 154, App |
| C 45 | 33 | 8.6 | 3282 | 1 | US-08-899-575-154 | Sequence 169, App |

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 INMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14
Query Match 14.1%; Score 54.2; DB 1; Length 7218;

Query Match 10.1%; Score 39; DB 1; Length 2277;
Best Local Similarity 45.0%; Pred. No. 0.37; Indels 0; Gaps 0;
Matches 147; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 14 CAGACGGGGCTGTGTGCGAGCTCAGATCTGTGGCTGTGGCAGATGTGGGGAG 73
DB 406 CAGTTCCGGCGCGTGTGTGGAGTGAACATCCCCCGCAAGCCCGACGCGCAAGATCGCGGC 465

QY 74 TACTCTGTGTGTGTGGGGAGGAGAGACCTCTGCCACTCTCACCCTGAAGGCCCTGCCA 133
DB 466 CAGTTCCGGCGCGTGTGTGGAGTGAACATCCCCCGCAAGCCCGACGCGCAAGATCGCGGC 465

QY 74 TACTCTGTGTGTGTGGGGAGGAGAGACCTCTGCCACTCTCACCCTGAAGGCCCTGCCA 133
DB 466 CAGTTCCGGCGCGTGTGTGGAGTGAACATCCCCCGCAAGCCCGACGCGCAAGATCGCGGC 465

QY 134 GCCAAGTTCCAGAGGGTCTGAGGAATGAAGAGCCCTGGAAGGGCCACAGCCATGTTG 193
DB 526 ATGAAGGAGATCAAGGCGCCGACCGTGGCCCTGGCCCAAGGACAAAGTAC 585

QY 194 TGGTGTGAACCTGAGCAAGTGGCCCTGTGGAGTGGAGAGGGCCCGAGAACCTCAGA 253
DB 586 AAGGACACCCAGAGCGTGAAGAGAGGCGCCGAGGAGAGGACATGGAGGAGGAGACGAC 645

QY 254 GATGGGACAGATACATCTTGAAGAGAGGCGCCGAGGAGGAGGACATGGAGGAGGAGACGAC 705
DB 646 GAGAGCTGAGAGAGAGGCGCCGAGGAGGAGGACATGGAGGAGGAGGAGGAGGAGGAGGAC 705

QY 314 CTGGCCATGGCGGACGCGCGGGAGTAC 340
DB 706 GACGACGACGACGACGACGAGGAGGAC 732

RESULT 7

US-09-098-487-5
; Sequence 5, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/098,487
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-098-487-5

Query Match 10.1%; Score 39; DB 2; Length 2277;
Best Local Similarity 45.0%; Pred. No. 0.37; Indels 0; Gaps 0;
Matches 147; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 14 CAGACGGGGCTGTGTGCGAGCTCAGATCTGTGGCTGTGGCAGATGTGGGGAG 73
DB 406 CAGTTCCGGCGCGTGTGTGGAGTGAACATCCCCCGCAAGCCCGACGCGCAAGATCGCGGC 465

QY 74 TACTCTGTGTGTGTGGGGAGGAGAGACCTCTGCCACTCTCACCCTGAAGGCCCTGCCA 133
DB 466 CAGTTCCGGCGCGTGTGTGGAGTGAACATCCCCCGCAAGCCCGACGCGCAAGATCGCGGC 465

QY 134 GCCAAGTTCCAGAGGGTCTGAGGAATGAAGAGCCCTGGAAGGGCCACAGCCATGTTG 193
DB 526 ATGAAGGAGATCAAGGCGCCGACCGTGGCCCTGGCCCAAGGACAAAGTAC 585

QY 194 TGGTGTGAACCTGAGCAAGTGGCCCTGTGGAGTGGAGAGGGCCCGAGAACCTCAGA 253
DB 586 AAGGACACCCAGAGCGTGAAGAGAGGCGCCGAGGAGAGGACATGGAGGAGGAGACGAC 645

QY 254 GATGGGACAGATACATCTTGAAGAGAGGCGCCGAGGAGGAGGACATGGAGGAGGAGACGAC 705
DB 646 GAGAGCTGAGAGAGAGGCGCCGAGGAGGAGGACATGGAGGAGGAGGAGGAGGAGGAGGAC 705

QY 314 CTGGCCATGGCGGACGCGCGGGAGTAC 340
DB 706 GACGACGACGACGACGACGAGGAGGAC 732

RESULT 8

US-09-620-312D-709/c
; Sequence 709, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: NO. 6569662e1 Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: PF_FL_genes Version 1.0
; SEQ ID NO 709
; LENGTH: 707
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (339)..(698)
US-09-620-312D-709

Query Match 9.8%; Score 37.6; DB 4; Length 707;
Best Local Similarity 44.8%; Pred. No. 0.66;
Matches 145; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
QY 11 AGGCAGGACGGGCTGTGTGGAGTGCAGATCTGTGGCTGTGGCAGATGTGGG 70

Db 379 AGATATGCTGGAAGTCTTCGATCTTGTCTCATCTTACCAATGCTGCTGAGTGTGG 320
QY 71 GAGTACTCTCTGTGTGTGGGAGGAGACCTCTGCCACTCTCACCGTGAAGGCCCTG 130
Db 319 GTGAATCTCTGGGAGCCCGAGGAGGACTCCACAGGATGGTCCGACGGGAACACCGG 260
QY 131 CCAGCCAAAGTTACAGAGGCTCTGAGGAATGAAGAGCCCGTGAAGGGGCCACAGCCATG 190
Db 259 AAGACACAGGTAAAGGCCGTGTTGGCAGCGGTGGTGTGGCAGGCTCCACACACTGG 200
QY 191 TTGTGGTGTGACTGACACAGGTGGCCCTGTGGAGTGGAGGAGGGCCCGAGAACCTC 250
Db 199 TGATGTCTTCGGCAGAAAGACCTTCACGACAGAGTCTAGGGCTAGTCTTCAGCAGCCCG 140
QY 251 AGAGATGGGACAGATACATCTCTGAGGACGAGGGGACCAAGTGTGAGCTGCAAGATCTGT 310
Db 139 CGACAGCGGCCCTGTCCGCCACGCTGGGCTTGGCTCCGCTGGGCTGGCGACGCGCGA 80
QY 311 GGCCTGCCATGGCGGACGCGGG 334
Db 79 AGCGCGCCCGGTGGGGCTCGGG 56

RESULT 9

US-09-308-345A-6
; Sequence 6, Application US/09308345A
; Patent No. 659665

; GENERAL INFORMATION:
; APPLICANT: BOEHM, Thomas;

; APPLICANT: DEAR, Neil T.
; TITLE OF INVENTION: No. 659665el calpains, their preparation and use

; FILE REFERENCE: 0050/47576
; CURRENT APPLICATION NUMBER: US/09/308,345A
; CURRENT FILING DATE: 1999-05-19

; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: WordPerfect v. 6.1

; SEQ ID NO 6
; LENGTH: 1975.

; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: CDS

; LOCATION: 44 . . 1963
US-09-308-345A-6

Query Match 9.7%; Score 37.4; DB 4; Length 1975;

Best Local Similarity 52.2%; Pred. No. 0.93;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 227 TGGAGGAAGGGCCCGAGAACTCAGAGATGGGACAGATACATCTCTGAGGAGGGG 286

Db 899 TGGGCGAGCGGAGTGGAAACGGCCCTGGAGTGACACCTCGAGGAGTGCAGAAAGTG 958

QY 287 ACCAGGTGTGAGTCAGATCTGTGGCTGGCCATGGCGACCGCGGGAGTACTTTGT 346

Db 959 AGCAAGTGTAGCGGGAGAGATGGTGTGACCGTGCAGGACGAGTGTCTGTGATG 1018

QY 347 GTGTGCGGCGAGGAGACCTCAGCCACGCTCACCATC 385

Db 1019 ACCTTCGAGGAGTGTGCCGTACTTACGGACATCATC 1057

RESULT 10

US-09-163-748C-3

; Sequence 3, Application US/09163748C
; Patent No. 6509172

; GENERAL INFORMATION:

; APPLICANT: DeBacker, Oliver
; APPLICANT: Van den Eynde, Benoit

; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode Member Of The Gage

; FAMILY OF INVENTION: Family, The Proteins Encoded, And Uses Thereof
; FILE REFERENCE: LUD 5558

; CURRENT APPLICATION NUMBER: US/09/163,748C
; CURRENT FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 27
; SEQ ID NO 3
; LENGTH: 9531

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: unsure

; LOCATION: 92,232,1041,7412,9038-9040

; OTHER INFORMATION: identity of several nucleotides not known
US-09-163-748C-3

Query Match 9.7%; Score 37.2; DB 4; Length 9531;

Best Local Similarity 51.2%; Pred. No. 1.5;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 186 CCATCTTGTGTGTAACGTGACGACAGTGGCCCTGTGGAGTGGAGGAGGGGCCGAGA 245

Db 2772 CCAGCCGAGCAGTTCAGTGTGAAGTGAACACGACACCTGAAGAGGGGAACACG 2831

QY 246 ACCTCAGAGATGGGACAGATACATCTCTGAGGAGGAGGGACCAAGTGTGAGCTGAGA 305

Db 2832 AACTCAACGTCAAGATCTCTGACGTCTCAGGAGGAGGAGGATGAGGAGCATCTGCAGG 2891

QY 306 TCTGTGGCTGCCATGGCGGACGCGGGGAGTACTTCTGTGTGCGGG 355

Db 2892 TCAAGTGAAGGAGGAGGAAACGCTGCTGTGTGCGGTGTGTG 2941

RESULT 11

US-08-669-161A-29

; Sequence 29, Application US/08669161A

; Patent No. 6013481

; GENERAL INFORMATION:

; APPLICANT: DeBacker, Olivier; Van den Eynde,
; APPLICANT: Benoit; Boon-Falleur, Thierry

; TITLE OF INVENTION: Isolated, Truncated Nucleic Acid
; TITLE OF INVENTION: Molecules Which Code For GAGE Tumor Rejection Antigen,

; TITLE OF INVENTION: The Tumor Rejection Antigen, and Uses Thereof
; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Felfe & Lynch

; STREET: 805 Third Avenue

; CITY: New York City

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage

; COMPUTER: IBM PS/2

; OPERATING SYSTEM: PC-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/669,161A

; FILING DATE: 24-June-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/531,662

; FILING DATE: 21-September-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/370,648

; FILING DATE: 10-January-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/250,162

; FILING DATE: 27-May-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/096,039

; FILING DATE: 22-July-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 6013481man D.

; REGISTRATION NUMBER: 30,946

; REFERENCE/DOCKET NUMBER: LUD 5443

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11461 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-669-161A-29

Query Match 9.7%; Score 37.2; DB 3; Length 11461;
Best Local Similarity 51.2%; Pred. No. 1.5;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 186 CCATGTTGTGTGAACAGCAAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAGA 245
Db 2772 CCACCCGAGCAGTTCAGTGTGATGAAGTGAACAGCAACACCTGAAGAGGGGAACGAC 2831
QY 246 ACCTCAGAGATGGGACAGATACATCTGTAGGAGGAGGGGACCAAGTGTGAGCTGCAGA 305
Db 2832 AACTCAACGTGAGATCTCTGAGCTGTCTCAGGAGGGAGAGGATCAGGGAGCATCTGCAG 2891
QY 306 TCTGTGCGCTGGCCATGCGGACGCGCGGGAGTACTTGTGTGTGCGGG 355
Db 2892 TCAAGGTGAGGAAAGGAAAGAACGCTCTGCTGTGTGTGCTGTGTG 2941

RESULT 12
US-09-687-050-5
; Sequence 5, Application US/09687050
; Patent No. 6632637
; GENERAL INFORMATION:
; APPLICANT: McGrew, Jeffrey T.
; TITLE OF INVENTION: VECTORS AND METHODS FOR RECOMBINANT PROTEIN EXPRESSION
; FILE REFERENCE: 2902-A
; CURRENT APPLICATION NUMBER: US/09/687,050
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: 60/159,177
; PRIOR FILING DATE: 1993-10-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 2478
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2478)
; NAME/KEY: mat_peptide
; LOCATION: (76)...()
; NAME/KEY: sig_peptide
; LOCATION: (1)...(75)
US-09-687-050-5

Query Match 9.6%; Score 37; DB 4; Length 2478;
Best Local Similarity 49.2%; Pred. No. 1.2;
Matches 97; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
QY 132 CAGCAAGTTCACAGAGGCTCTGAGGAATGAGAGGCGGTGGAGGGCCACAGCCATGT 191
Db 1016 CAGCATGTCGCCAGTGGAGATCAGCAAGACATCTCTGCCAGAGAGATCAGCGTG 1075
QY 192 TGTGTTGTAACAGCAAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAGAACTCA 251
Db 1076 TCGATGTGTGGAGTTGTTTGAAGCCCGGTGGAGTGTGAGGAGGAGGAGGTAGAG 1135
QY 252 GAGATGGGACAGATACATCTGTAGGAGGAGGGGACCAAGTGTGAGCTGCAGATCTGTG 311
Db 1136 AAGAAAAGGAGGCTTCTGTGCATCGCCCTGAGAGCAGCGGATGACTTCCAGAGGGAA 1195
QY 312 GCCTGGCCATGGCGAC 328

Db 1196 GGGAGGCAATTGTGGCC 1212

RESULT 13
US-09-921-667-15
; Sequence 15, Application US/09921667
; Patent No. 6652854
; GENERAL INFORMATION:
; APPLICANT: Mohler, Kendall M.
; APPLICANT: Barone, Dauphine S.
; APPLICANT: Peschon, Jacques J.
; APPLICANT: Kennedy, Mary K.
; APPLICANT: Pluennake, John D.
; TITLE OF INVENTION: METHODS FOR TREATING AUTOIMMUNE AND CHRONIC INFLAMMATORY CONDITIONS
; FILE REFERENCE: 2959-A
; CURRENT APPLICATION NUMBER: US/09/921,667
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 60/224,079
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15
; LENGTH: 2478
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2475)
; OTHER INFORMATION:
; NAME/KEY: mat_peptide
; LOCATION: (76)...()
; OTHER INFORMATION:
US-09-921-667-15

Query Match 9.6%; Score 37; DB 4; Length 2478;
Best Local Similarity 49.2%; Pred. No. 1.2;
Matches 97; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
QY 132 CAGCAAGTTCACAGAGGCTCTGAGGAATGAGAGGCGGTGGAGGGCCACAGCCATGT 191
Db 1016 CAGCATGTCGCCAGTGGAGATCAGCAAGACATCTCTGCCAGAGAGATCAGCGTG 1075
QY 192 TGTGTTGTAACAGCAAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAGAACTCA 251
Db 1076 TCGATGTGTGGAGTTGTTTGAAGCCCGGTGGAGTGTGAGGAGGAGGAGGTAGAG 1135
QY 252 GAGATGGGACAGATACATCTGTAGGAGGAGGGGACCAAGTGTGAGCTGCAGATCTGTG 311
Db 1136 AAGAAAAGGAGGCTTCTGTGCATCGCCCTGAGAGCAGCGGATGACTTCCAGAGGGAA 1195
QY 312 GCCTGGCCATGGCGAC 328
Db 1196 GGGAGGCAATTGTGGCC 1212

RESULT 14
US-09-016-434-1319
; Sequence 1319, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 845-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1314:
SEQUENCE CHARACTERISTICS:
LENGTH: 3597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g33833
US-09-016-434-1319

Query Match          9.6%; Score 37; DB 4; Length 3597;
Best Local Similarity 49.2%; Pred. No. 1.3; Mismatches 100; Indels 0; Gaps 0;
Matches 97; Conservative 0;

QY 132 CAGCCAAAGTTTACAGAGGGTCTGAGGAATGAAGAGGCCGCTGGAAGGGCCGACAGCCATGT 191
Db 1191 CAGCATGGTCCAGTGGAGATCAGCAAGACAGTCTCTGGCCAGAGAGCATCAGCGTGG 1250

QY 192 TGTGGTGAAGTGAAGAGTGGCCCTCTGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 251
Db 1251 TCGCATGTGTGGAGTTGTTTGGAGCCCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1310

QY 252 GAGATGGGGACAGATACATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 311
Db 1311 AAGAAAAGGAGCTTCTGTGCATCGCTGAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGG 1370

QY 312 GCCTGGCCATGGCGGAC 328
Db 1371 GGGAGGGCATTTGGGCC 1387

Search completed: September 19, 2004, 22:12:36
Job time : 35.5752 secs
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 845-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1319:
SEQUENCE CHARACTERISTICS:
LENGTH: 3597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g33833
US-09-016-434-1319

Query Match          9.6%; Score 37; DB 4; Length 3597;
Best Local Similarity 49.2%; Pred. No. 1.3; Mismatches 100; Indels 0; Gaps 0;
Matches 97; Conservative 0;

QY 132 CAGCCAAAGTTTACAGAGGGTCTGAGGAATGAAGAGGCCGCTGGAAGGGCCGACAGCCATGT 191
Db 1191 CAGCATGGTCCAGTGGAGATCAGCAAGACAGTCTCTGGCCAGAGAGCATCAGCGTGG 1250

QY 192 TGTGGTGAAGTGAAGAGTGGCCCTCTGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 251
Db 1251 TCGCATGTGTGGAGTTGTTTGGAGCCCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1310

QY 252 GAGATGGGGACAGATACATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 311
Db 1311 AAGAAAAGGAGCTTCTGTGCATCGCTGAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGG 1370

QY 312 GCCTGGCCATGGCGGAC 328
Db 1371 GGGAGGGCATTTGGGCC 1387
```

```
RESULT 15
US-09-023-655-1314
Sequence 1314, Application US/09023655
Patent No. 6607879
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
APPLICANT: Susan G. Stuart
APPLICANT: Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
NUMBER OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSES: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 12:53:28 ; Search time 158.742 Seconds
(without alignments)
10303.209 Million cell updates/sec

Title: US-10-077-130-4_COPY_10286_10670

Perfect score: 385
Sequence: 1 acatagctgagcaggagc.....cctcagccacgctcaccatc 385

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04.*

- 1: geneseq1980s.*
- 2: geneseq1990s.*
- 3: geneseq2000s.*
- 4: geneseq2001as.*
- 5: geneseq2001bs.*
- 6: geneseq2002s.*
- 7: geneseq2003as.*
- 8: geneseq2003bs.*
- 9: geneseq2003cs.*
- 10: geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|---------------------|
| 1 | 385 | 100.0 | 7564 | 7 | ABZ24581 Human cel |
| 2 | 385 | 100.0 | 14061 | 6 | ABV99363 Human NOV |
| 3 | 385 | 100.0 | 14109 | 6 | ABV99362 Human NOV |
| 4 | 385 | 100.0 | 24120 | 7 | ABX11642 Human ser |
| 5 | 253.6 | 65.9 | 352 | 5 | AAF64470 Novel hum |
| 6 | 202 | 52.5 | 898 | 6 | ABL58139 Human pho |
| 7 | 195.4 | 50.8 | 642 | 6 | ABQ61169 Obscurin |
| 8 | 195.4 | 50.8 | 707 | 6 | ABQ61180 Obscurin |
| 9 | 190.6 | 49.5 | 2155 | 6 | ABK99965 DNA encod |
| 10 | 71.6 | 18.6 | 1005 | 6 | ABN21414 Human ORF |
| 11 | 47.6 | 12.4 | 2737 | 4 | AAI59251 Human pol |
| 12 | 47.6 | 12.4 | 3399 | 4 | AAI61037 Human pol |
| 13 | 47.6 | 12.4 | 5382 | 9 | ADD14722 Human src |
| 14 | 46.8 | 12.2 | 81940 | 4 | AAK505390 Human tit |
| 15 | 46.8 | 12.2 | 81940 | 6 | ABK64829 Human ben |
| 16 | 46.8 | 12.2 | 93801 | 8 | ABX13540 Human RGS |
| 17 | 46.2 | 12.0 | 60 | 6 | ABN40768 Human spl |
| 18 | 43.4 | 11.3 | 1845 | 5 | AAI24162 Human sec |
| 19 | 42.8 | 11.1 | 2254 | 3 | AAK93403 Human sec |
| 20 | 42.8 | 11.1 | 2254 | 5 | AAI24183 Human sec |
| 21 | 41.8 | 10.9 | 716 | 6 | ABK35726 CDNA sequ |
| 22 | 41.8 | 10.9 | 1948 | 2 | ABV74346 Human IL- |
| 23 | 41.8 | 10.9 | 1949 | 2 | ABV74347 Human IL- |

| | | | | | | |
|----|------|------|-------|----|----------|------------|
| 24 | 41.8 | 10.9 | 2170 | 6 | ABQ54970 | Human ova |
| 25 | 40.6 | 10.5 | 3761 | 4 | ABL05157 | Drosophila |
| 26 | 40.6 | 10.5 | 6069 | 9 | ADD18703 | Human dis |
| 27 | 40.6 | 10.5 | 7066 | 10 | ADE77197 | Human gen |
| 28 | 39.4 | 10.2 | 1359 | 2 | AAZ17254 | Human sapi |
| 29 | 39 | 10.1 | 2277 | 2 | AAV13836 | Human sapi |
| 30 | 39 | 10.1 | 2277 | 2 | AAV05372 | Human sapi |
| 31 | 38.6 | 10.0 | 5452 | 9 | ADC86736 | Human sapi |
| 32 | 38.2 | 9.9 | 1251 | 9 | ADC30509 | Human sapi |
| 33 | 38.2 | 9.9 | 2488 | 4 | AAH18453 | Human sapi |
| 34 | 38.2 | 9.9 | 67251 | 9 | ADC26995 | Sorangium |
| 35 | 38 | 9.9 | 8047 | 4 | ABA07219 | Human sapi |
| 36 | 38 | 9.9 | 8047 | 4 | AAK89856 | Human sapi |
| 37 | 37.8 | 9.8 | 17127 | 7 | ADA98892 | Human sapi |
| 38 | 37.8 | 9.8 | 17127 | 7 | ADA44493 | Human sapi |
| 39 | 37.8 | 9.8 | 17127 | 9 | ADC20927 | Human sapi |
| 40 | 37.6 | 9.8 | 672 | 6 | ABQ98892 | Human sapi |
| 41 | 37.6 | 9.8 | 707 | 4 | AAI58818 | Human sapi |
| 42 | 37.6 | 9.8 | 707 | 8 | ADB48799 | Novel hum |
| 43 | 37.6 | 9.8 | 1068 | 7 | ADA70882 | Rice gene |
| 44 | 37.6 | 9.8 | 2121 | 4 | AAK94576 | Human sapi |
| 45 | 37.6 | 9.8 | 2437 | 6 | ABN85621 | Human sapi |

ALIGNMENTS

RESULT 1

ABZ24581
ID ABZ24581 standard; cDNA; 7564 BP.

XX AC ABZ24581;

XX 31-MAR-2003 (first entry)

XX Human cell adhesion and extracellular matrix protein 4 cDNA.

XX Cell adhesion and extracellular matrix protein 4; CADECM-4; human; anti-HIV; virucide; antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nootropic; anticonvulsant; antifertility; antitartaroclastic; antiaesthetic; immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic; antitumor; thyromimetic; neuroprotective; osteoplastic; antiparastic; antihelminthic; antiparasitic; uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial; protozoacide; fungicide; gynaecological; titin; gene therapy; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
XX CDS 241..7227
XX /*tag= a
XX /*product= "Human CADECM-4"

XX WC200288322-A2.

XX 07-NOV-2002.

XX 01-MAY-2002; 2002WO-US013874.

XX 02-MAY-2001; 2001US-0288290P.

XX 15-MAY-2001; 2001US-0292468P.

XX 28-JUN-2001; 2001US-0298616P.

XX 04-JAN-2002; 2001US-0301672P.

XX 04-JAN-2002; 2002US-0345008P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Lee EA, Duggan BM, Thangavelu K, Honchell CD, Ding Li, Tran UK; Hillman JL, Baughn MR, Kallick DA, Lee S, Warren BA, Xu Y, Tran UK; Lal PG, Thornton M, Hafalia AUA, Yao MG, Nguyen DB, Gandhi AR; Khan FA, Walhia NK, Griffin JA, Chinn AM, Elliott VS, Ramkumar J; Arvizu CS, Forsythe LJ;

| | |
|----|---|
| XX | 15-FEB-2002; 2002US-00077130. |
| PF | (MILL-) MILLENNIUM PHARM INC. |
| XX | Kapeller-Libermann R, Acton SL; |
| XX | WPI; 2003-298729/29. |
| DR | P-PSDB; ABG76187. |
| XX | Novel isolated human protein kinase, designated 59079 or 12599 |
| PT | polypeptide, useful as diagnostic and therapeutic agents for preventing |
| PT | cardiovascular diseases, proliferative disorders, and protein kinase |
| PT | disorders. |
| XX | Claim 2; Page 58-84; 119pp; English. |
| PS | The invention relates to an isolated human serine/threonine or protein |
| XX | kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule |
| CC | comprising at least 8% identity to the nucleic acids appearing as |
| CC | ABX11641 and ABX11642 or their complement, a naturally occurring variant |
| CC | of the kinases or their fragments. Also included are a non-human host |
| CC | cell containing the nucleic acids, an antibody specific for the proteins, |
| CC | identifying a compound which binds to the kinase (by contacting the |
| CC | kinase or a cell expressing the kinase with a test compound and |
| CC | determining whether the kinase binds to the test compound) and modulating |
| CC | the activity of kinase using the identified compound. The kinases and |
| CC | their encoding nucleic acids are useful as diagnostic and therapeutic |
| CC | agents for preventing a disease or condition associated with an aberrant |
| CC | or unwanted 59079 or 12599 activity in a subject, including |
| CC | cardiovascular diseases such as heart failure, and myocardial infarction; |
| CC | disorders involving blood vessels such as atherosclerosis, and Kaposi's |
| CC | sarcoma; blood platelets disorder such as thrombocytopoenia, leukaemia, |
| CC | Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders |
| CC | such as cancer; and protein kinase disorders such as autoimmune |
| CC | rheumatoid arthritis, and multiple sclerosis (many examples of diseases |
| CC | and disorders are included in the specification). The kinases, their |
| CC | encoding nucleic acids and antibodies are useful in screening assays, |
| CC | detection assays (e.g. forensic biology), and predictive medicine (e.g. |
| CC | diagnostic assays, prognostic assays, and monitoring clinical trials and |
| CC | pharmacogenomics). The kinases and their encoding nucleic acids are |
| CC | useful as query sequences to perform a search against public databases to |
| CC | identify other family members or related sequences. The present sequence |
| CC | encodes the kinase 12599 |
| XX | Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other; |
| SQ | Query Match 100.0%; Score 385; DB 7; Length 24120; |
| | Best Local Similarity 100.0%; Pred. NO. 4.7e-86; |
| | Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| QY | 1 ACATAGCCTGAGCGAGGACGGGGCTGTGCCAGCTGCAGATCTGTGCCTGGCTGTGGC 60 |
| Db | 10286 ACATAGCCTGAGCAGGACGGGGCTGTGCCAGCTGCAGATCTGTGCCTGGCTGTGGC 10345 |
| QY | 61 AGATGTGGGGAGTACTCTGTGTGTGTGGGGAGGAGGACCTCTGCCACTCTCACCGT 120 |
| Db | 10346 AGATGTGGGGAGTACTCTGTGTGTGTGGGGAGGAGGACCTCTGCCACTCTCACCGT 10405 |
| QY | 121 GAAGGCCCTGCCAGCCAAAGTTACAGAGGGTCTGAGGAATGAGAGGCCGTGGAAGGGC 180 |
| Db | 10406 GAAGGCCCTGCCAGCCAAAGTTACAGAGGGTCTGAGGAATGAGAGGCCGTGGAAGGGC 10465 |
| QY | 181 CACAGCCCATGTTGTGGTGTTGAATTCAGCAAGGTGGCCCCCTGTGGAGTGGAGAGGGGCC 240 |
| Db | 10466 CACAGCCCATGTTGTGGTGTTGAATTCAGCAAGGTGGCCCCCTGTGGAGTGGAGAGGGGCC 10525 |
| QY | 241 CGAAGAACCCTCAGAGATGGGGAAGATACATCTCTGAGGAGGAGGGGAACAAGTGTAGCT 300 |
| Db | 10526 CGAAGAACCCTCAGAGATGGGGAAGATACATCTCTGAGGAGGAGGGGAACAAGTGTAGCT 10585 |

| | | | |
|----|-------|--|-------|
| Qy | 301 | GCAGATCTGTGGCGCTGGCCATCGCGGAGCGCCGGGAGTACTGTGTGTGTGCGGGCAGGA | 360 |
| | | | |
| Db | 10586 | GCAGATCTGTGGCGCTGGCCATCGCGGAGCGCCGGGAGTACTGTGTGTGTGCGGGCAGGA | 10645 |
| | | | |
| Qy | 361 | GAGGACCTCAGCGACGCTCACCATC | 385 |
| | | | |
| Db | 10646 | GAGGACCTCAGCGACGCTCACCATC | 10670 |

| | |
|------------|--|
| RESULT 5 | |
| AAF64470/c | |
| ID | AAF64470 standard; cDNA; 352 BP. |
| XX | |
| XX | AAF64470; |
| XX | |
| DT | 09-APR-2001 (first entry) |
| XX | |
| DE | Novel human polynucleotide, SEQ ID NO: 226. |
| XX | |
| KW | Human; cytostatic; gene therapy; colon cancer; prostate cancer; |
| KW | breast cancer; lung cancer; cancer detection; ss. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| FN | WO200102568-A2. |
| XX | |
| PD | 11-JAN-2001. |
| XX | |
| FF | 30-JUN-2000; 20COWO-US01-8374. |
| XX | |
| PR | 02-JUL-1999; 99US-0142310P. |
| XX | |
| XX | 02-JUL-1999; 99US-0142311P. |
| XX | |
| PA | (CHIR) CHIRON CORP. |
| PA | (HYSE-) HYSEQ INC. |
| XX | |
| PI | Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J; |
| PI | Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G; |
| PI | Drmanac R, Crkjenjakov R, Drmanac S, Dickson M, Labat I; |
| PI | Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B; |
| XX | |
| DR | WPI; 2001-091805/10. |
| XX | |
| PT | Library of polynucleotides for diagnosing a cancerous state of a |
| PT | mammalian cell and detecting cancer, particularly of the colon or |
| PT | prostate, comprises 3351 human polynucleotide sequences. |
| XX | |
| PS | Claim 9; Page 576; 1046pp; English. |
| XX | |
| CC | The present sequence is one of 3351 sequences in a library of human |
| CC | polynucleotides. The library is used to detect differentially expressed |
| CC | genes correlated with a cancerous state of a mammalian cell and can |
| CC | detect colon, prostate, breast and lung cancer. The library can be used |
| CC | to produce probes for detection of mRNA and to produce additional copies |
| CC | of the polynucleotides. The probes can be used for chromosome mapping of |
| CC | the polynucleotide and for detection of transcription levels. Ribozymes |
| CC | or antisense oligonucleotides can be generated. The polynucleotides and |
| CC | their gene products are used as genetic or biochemical markers (e.g. in |
| CC | blood or tissues) that will detect the earliest changes along the |
| CC | carcinogenesis pathway and/or monitor the efficacy of therapies and |
| CC | preventive interventions. The polynucleotides, polypeptides and |
| CC | antibodies against them can be used in pharmaceutical compositions to |
| CC | treat the cancers and proliferative disorders such as neoplasia, |
| CC | dysplasia and hyperplasia |
| XX | |
| SQ | Sequence 352 BP; 56 A; 138 C; 84 G; 73 T; 0 U; 1 Other; |
| | |
| | Query Match 65.9%; Score 253.6; DB 5; Length 352; |
| | Best Local Similarity 86.4%; Fred. No. 1.1e-53; |
| | Matches 280; Conservative 0; Mismatches 44; Indels 0; Gaps 0; |
| QY | 62 GATGCTGGGAGTACTCTGTGTGTGTGGGAGGAGGACCTGTGCCACTCTCACCGT 121 |

Db 342 GACGCTGGGAGTACCTGTGCATGTGCGGGAAGGAGGAGGACCTCAGCCATGCTCACCGTC 283
QY 122 AAGCCCTGCCAGCAAGTTTCACAGAGGGTCTGAGGAATGAAGAGCCGTGGAAGGGGCC 181
Db 282 AGGCCCATGCTTCCAAAGTTTCATAGAGGGTCTGAGGAATGAAGAGGCCACAGAAAGGGGAC 223
QY 182 ACAGCCATGTTGTGCTGTGAACCTCAGCAAGGTGGCCCTCTGAGAGTGGAGAAAGGGGCC 241
Db 222 ACGCCACCGCTGTGCTGTGAGCTGAGCAAGCGGCCACCGGTGAGTGGAGAAAGGGGCAT 163
QY 242 GAGAACCTCAGAGATGGGGACAGATACATCTGTAGGCAGAGGGGACCAAGGTGTGAGCTG 301
Db 162 GAGACCCCTCAGAGATGGGGACAGACACAGCTGTAGGCAGGATGGGTCCAGGTGTGAGCTG 103
QY 302 CAGATCTGTGCCCTGGCCATGGCGGAGCGCGGGGAGTACTTGTGTGTCGGGCGAGGAG 361
Db 102 CAGATCGGTGGCTTGGCTGTGGTGGATGCCGGGAGTACTCTGTGTGTCTCGGCGAGGAG 43
QY 362 AGGACCTCAGCCACGCTCACCATC 385
Db 42 AGGACCTCAGCCACACTCAGCTGTC 19

RESULT 6
ABL58139
ID ABL58139 standard; cDNA; 898 BP.
XX AC ABL58139;
XX DT 30-JUL-2002 (first entry)
XX DE Human phosphorylcholine/ethanolamine transferase 15.29 coding sequence.
XX KW Human; phosphorylcholine/ethanolamine transferase 15.29; enzyme;
XX KW phosphorylcholine; ethanolamine; transferase; tumour; inflammation;
XX KW cystostatic; anti-inflammatory; gene; gene therapy; ss.
XX OS Homo sapiens.
XX Key Location/Qualifiers
XX FT 411..830
XX CDS /*tag= a
XX FT /product= "Human phosphorylcholine/ethanolamine
XX FT transferase 15.29"
XX CN1331315-A.
XX PN 16-JAN-2002.
XX PD 30-JUN-2000; 2000CN-00116906.
XX PF 30-JUN-2000; 2000CN-00116906.
XX PR (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX PA Mao Y, Xie Y;
XX PI WPI; 2002-292894/34.
XX DR P-PSDB; ABB83109.
XX PT Human phosphorylcholine/ethanolamine transferase 15.29 and encoding
XX PT polynucleotide useful for treating tumors and inflammation.
XX PS Claim 6; Page 27 (Disclosure); 34pp; Chinese.
XX CC The present sequence is the coding sequence for human phosphorylcholine/
XX CC ethanolamine transferase 15.29. The enzyme and its coding sequence are
XX CC useful for treating diseases such as tumours and inflammation
XX SQ Sequence 898 BP; 169 A; 245 C; 239 G; 245 T; 0 U; 0 Other;
Query Match 52.5%; Score 202; DB 6; Length 898;

Best Local Similarity 83.6%; Pred. No. 9.3e-41;
Matches 229; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 112 TCTACCTGTGAAGCCCTGCCAGCAAGTTTCACAGAGGGTCTGAGGAATGAAGAGGCCGT 171
Db 335 TCTGACCTCCCGAGCCCTGCCCGCAAGTTTCACAAAGGGTCTGAGGAATGAAGAGGCCAC 394
QY 172 GGAAGGGGCCACAGCCATGTTGTGTTGTGAACCTCAGCAAGGTGGCCCTCTGTGAGTGGAG 231
Db 395 AGAAGGGGCCACGACTATGTTGTCAGTGTGAGCTGAGCAAGGTGGCCCTCTGTGAGTGGAG 454
QY 232 GAAGGGGCCACAGAACCTCAGAGATGGGGACAGATACATCTGTAGGCAGAGGGGACAG 291
Db 455 GAAGGGACCTGAAACCCCTCAGAGATGGGGACAGATACAACTGAGGAGGATGGGACAG 514
QY 292 GTGTGAGCTGAGATCTGTGGCTGGCCATGGCGGAGCCCGGGGAGTACTTGTGTGTGTG 351
Db 515 ATGTGAGCTGAGATTCATGGCTGTCCGTGGCAGACACTGGGGAGTACTCATGTGTATG 574
QY 352 CGGGCAGAGAGGAGCCTCAGCAAGCTCAGCCATCACCATC 385
Db 575 TGGTCAGGAGAAGACGTCGGCCACTCTCAGCTGC 608
RESULT 7
ABO61169
ID ABO61169 standard; cDNA; 642 BP.
XX AC ABO61169;
XX DT 26-FEB-2003 (first entry)
XX DE Obscure OBSCN gene encoding sequence.
XX KW Neuroprotective; immunomodulator; cancer; cytostatic; anti-inflammatory;
XX KW gene therapy; nutritional supplement; wound; burn; ulcer;
XX KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
XX KW autoimmune disorder; inflammation; vulnerability; gene; ss.
XX OS Homo sapiens.
XX PN WO200231111-A2.
XX PD 18-APR-2002.
XX PF 11-OCT-2001; 2001WO-US027760.
XX PR 12-OCT-2000; 2000US-00687527.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX DR WPI; 2002-426278/45.
XX DR N-PSDB; ABP43925.
XX PT New polypeptides and their encoded proteins, useful as nutritional
XX PT sources or supplements, or in gene therapy, particularly for treating
XX PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
XX PT inflammation.
XX PS Claim 1; SEQ ID # 382; 357pp + Sequence Listing; English.
XX CC The invention relates to 446 newly isolated polynucleotide sequences. The
XX CC activity of polynucleotides of the invention may be described as,
XX CC vulnary, neuroprotective, immunomodulator, cytostatic and anti-
XX CC inflammatory. Compositions comprising nucleic acids of the invention are
XX CC useful for treating a mammalian subject, or as nutritional sources or
XX CC supplements. These are useful in gene therapy, particularly for treating
XX CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
XX CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
XX CC inflammation. The nucleic acids and polypeptides are also useful in

CC diagnostic and research methods. The sequences given in records ABQ60788-
CC ABQ61233 represent polynucleotides of the invention. NOTE: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 642 BP; 147 A; 169 C; 205 G; 121 T; 0 U; 0 Other;

Query Match 50.8%; Score 195.4; DB 6; Length 642;
Best Local Similarity 84.3%; Pred. No. 3.8e-39;
Matches 220; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 125 GCCTGCCAGCAAGTTCACAGAGGGTCTGAGGAATCAAGAGGCCGTGGAGGGGCCACCA 184
DB 227 GCTTACCATCAAGTTCACAGAGGGTCTGAGGAACGAGAGGCCACAGAGGGGCCAACA 286
QY 185 GCCATGTTGTGTGTAACCTGAGCAAGTGGCCCTCTGTGAGTGGAGAGAGGGCCCGAG 244
DB 287 GCCGTGTGCGGTGTGAGCTGAGCAAGATGGCCCTCTGTGAGTGGAGAGGGGCCATGAG 346
QY 245 AACCTCAGAGATGGGACAGATACATCTGAGGCAGGAGGGACCGAGTGTGAGCTGCAG 304
DB 347 ACCCTCAGAGATGGGACAGACAGACCTGAGGCAGGACCGGGCCAGTGTGAGCTGCAG 406
QY 305 ATCTGTGCGCTGGCCATGGCGAGCGCCGGGAGTACTTGTGTGTGTCGGGCGAGGAGG 364
DB 407 ATCCGCGCTCTGTGGCAGAGGACGCTGGGAGTACCTGTGATGTGCGGGAAGGAGG 466
QY 365 ACCTCAGCCAGCTCACCATC 385
DB 467 ACCTCAGCCATGCTCACCGTC 487

RESULT 8
ABQ61180
ID ABQ61180 standard; cDNA; 707 BP.
XX
AC ABQ61180;
XX
DT 26-FEB-2003 (first entry)
XX
DE Obscurin OBSCN gene #2 encoding sequence.

XX Neuroprotective; immunomodulator; cancer; cytostatic; anti-inflammatory;
KW gene therapy; nutritional supplement; wound; burn; ulcer;
KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW autoimmune disorder; inflammation; vulnery; gene; ss.

OS Homo sapiens.
XX
PN WO200231111-A2.
XX
PD 18-APR-2002.
XX
PF 11-OCT-2001; 2001WO-US027760.
XX
PR 12-OCT-2000; 2000US-00687527.
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
XX WPI; 2002-426278/45.
DR N-PSDB; ABP43936.
XX

XX New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation.

PS Claim 1; SEQ ID # 393; 357pp + Sequence Listing; English.
XX

CC The invention relates to 446 newly isolated polynucleotide sequences. The
CC activity of polynucleotides of the invention may be described as,
CC vulnery, neuroprotective, immunomodulator, cytostatic and anti-
CC inflammatory. Compositions comprising nucleic acids of the invention are
CC useful for treating a mammalian subject, or as nutritional sources or
CC supplements. These are useful in gene therapy, particularly for treating
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC inflammation. The nucleic acids and polypeptides are also useful in
CC diagnostic and research methods. The sequences given in records ABQ60788-
CC ABQ61233 represent polynucleotides of the invention. NOTE: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 707 BP; 154 A; 198 C; 213 G; 142 T; 0 U; 0 Other;

Query Match 50.8%; Score 195.4; DB 6; Length 707;
Best Local Similarity 84.3%; Pred. No. 3.9e-39;
Matches 220; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 125 GCCTGCCAGCAAGTTCACAGAGGGTCTGAGGAATCAAGAGGCCGTGGAGGGGCCACCA 184
DB 227 GCTTACCATCAAGTTCACAGAGGGTCTGAGGAACGAGAGGCCACAGAGGGGCCAACA 286
QY 185 GCCATGTTGTGTGTAACCTGAGCAAGTGGCCCTCTGTGAGTGGAGAGAGGGCCCGAG 244
DB 287 GCCGTGTGCGGTGTGAGCTGAGCAAGATGGCCCTCTGTGAGTGGAGAGGGGCCATGAG 346
QY 245 AACCTCAGAGATGGGACAGATACATCTGAGGCAGGAGGGACCGAGTGTGAGCTGCAG 304
DB 347 ACCCTCAGAGATGGGACAGACAGACCTGAGGCAGGACCGGGCCAGTGTGAGCTGCAG 406
QY 305 ATCTGTGCGCTGGCCATGGCGAGCGCCGGGAGTACTTGTGTGTGTCGGGCGAGGAGG 364
DB 407 ATCCGCGCTCTGTGGCAGAGGACGCTGGGAGTACCTGTGATGTGCGGGAAGGAGG 466
QY 365 ACCTCAGCCAGCTCACCATC 385
DB 467 ACCTCAGCCATGCTCACCGTC 487

RESULT 9
ABK99965
ID ABK99965 standard; DNA; 2155 BP.
XX
AC ABK99965;
XX
DT 21-OCT-2002 (first entry)
XX
DE DNA encoding human secreted protein SGP-48.

XX Secreted protein; SGP; human; cell proliferative disorder; cancer;
KW keratosis; arteriosclerosis; atherosclerosis; cirrhosis; hepatitis;
KW psoriasis; autoimmune disorder; inflammatory disorder; AIDS; arthritis;
KW acquired immunodeficiency syndrome; adult respiratory distress syndrome;
KW Addison's disease; allergy; asthma; osteoporosis; autoimmune thyroiditis;
KW Crohn's disease; dermatitis; diabetes; Graves' disease; haemodialysis;
KW glomerulonephritis; scleroderma; systemic lupus erythematosus; uveitis;
KW systemic sclerosis; ulcerative colitis; infection; trauma; Pick disease;
KW cardiovascular heart disease; congestive heart failure; angina; epilepsy;
KW hypertensive heart disease; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; amyotrophic lateral sclerosis; stroke; dementia;
KW Huntington's disease; multiple sclerosis; neuromuscular disorder;
KW metabolic disorder; endocrine disorder; coxix myopathy; mental disorder;
KW schizophrenic disorder; developmental disorder; anaemia; epilepsy;
KW hypothyroidism; glaucoma; sensorineural hearing loss; cataract;
KW transgenic animal; gene; ds.

OS Homo sapiens.
XX
PN WO200248337-A2.
XX

PD 20-JUN-2002.

XX 12-DEC-2001; 2001WO-US048517.

XX 13-DEC-2000; 2000US-02556339P.

PR 21-DEC-2000; 2000US-0257852P.

PR 05-JAN-2001; 2001US-0260105P.

PR 18-JAN-2001; 2001US-0282932P.

PR 18-JAN-2001; 2001US-0283096P.

PR 19-JAN-2001; 2001US-0263090P.

PR 02-FEB-2001; 2001US-0265926P.

XX (INCYTE GENOMICS INC.

XX Griffin JA, Yao MG, Duggan BM, Yue H, Ding L, Lal PG, Lee EA;

PI Ramkumar J, Thangavelu K, Xu Y, Lee S, Tang YT, Nguyen DB;

PI Warren BA, Honchell CD, Gietzen KJ, Baughn MR, Gandhi AR, Arvizu C;

PI Wallia NK, Lu Y, Elliott VW, Lu DAM, Hafalia AJA, Azimzai Y;

XX Khan FA, Tran UK;

XX WPI: 2002-583509/62.

DR P-PSDB; ABG69668.

XX Novel human secreted proteins and polynucleotides for diagnosing,

PT preventing or treating cell proliferative, autoimmune/inflammatory,

PT cardiovascular, neurological and developmental disorders.

XX Claim 5; Page 230; 234pp; English.

XX The invention describes an isolated polypeptide chosen from secreted

CC proteins (I), SECP 1-54. (I) and the polynucleotide encoding it (II) are

CC useful for screening a compound for effectiveness as an agonist or

CC antagonist of (I) or compound that alters expression of (II). (I), the

CC identified agonist and antagonist are useful for treating a disease or

CC condition associated altered expression of functional SECP in a patient.

CC An antibody specific to (I) is useful for detecting the presence of (I),

CC purifying (I) from a sample and for diagnosing a condition or disease

CC associated with expression of SECP in a subject or in a biological

CC sample. (I) and (II) and modulators of (I) are useful for diagnosis,

CC treatment and prevention of cell proliferative disorders (e.g. cancer,

CC keratosis, arteriosclerosis, atherosclerosis, cirrhosis, hepatitis and

CC psoriasis), autoimmune/inflammatory disorders (e.g. acquired

CC immunodeficiency syndrome (AIDS), adult respiratory distress syndrome,

CC Addison's disease, allergies, asthma, osteoporosis, autoimmune

CC thyroiditis, Crohn's disease, dermatitis, diabetes, Graves' disease,

CC glomerulonephritis, arthritis, scleroderma, systemic lupus erythematosus,

CC systemic sclerosis, ulcerative colitis, haemodialysis, uveitis; viral,

CC bacterial, fungal, parasitic, protozoal, helminthic infections and

CC trauma), cardiovascular disorders (e.g. congestive heart failure, angina,

CC hypertensive heart disease), neurological disorders (e.g. Alzheimer's and

CC Pick disease, Parkinson disease, amyotrophic lateral sclerosis, epilepsy,

CC stroke, Huntington's disease, multiple sclerosis, dementia, neuromuscular

CC disorders, metabolic, endocrine and toxic myopathies, mental disorders,

CC schizophrenic disorders, and developmental disorders (e.g. anaemia,

CC epilepsy, hypothyroidism, glaucoma, sensorineural hearing loss and

CC cataract). (II) is useful for creating transgenic animals to model human

CC disease and to detect and quantify gene expression in biopsied tissues in

CC which expression of SECP is correlated with disease. This sequence

CC encodes a human secreted protein (SCEP)

XX Query Match 49.5%; Score 190.6; DB 6; Length 2155;

XX Best Local Similarity 83.1%; Pred. No. 7.6e-39;

XX Matches 21; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 125 GCCTGCGACCAAGTTACAGAGGGTCTGAGGAATGAAGAGCCGTGGAAGGGCCACA 184

Db 205 GCCCTGCTGCGAGTTACAGAGGGTCTGAGGAATGAAGAGCCGTGGAAGGGCCACA 264

Qy 185 GCCATGTTGGTGTGAATGAGTGAAGTGGCCCTGTGAGTGGAGGAGGGCCCGAG 244

Db 265 GCCACATGCAATGTAGTGTGAGCAGAGGAGCCCTGTGAGTGGAGGAAGGCTTGAG 324

QY 245 AACCTCAGAGATGGGACAGATACATCTCTGAGCGAGGAGGGACACAGTGTGAGCTGCAG 304

Db 325 GCTCTCAGAGATGGGACAAATACAGCCTGAGACAGACGGGGCTGTGTGAGCTGCAG 384

QY 305 ATCTGTGGCTGGCCATGGCGAGCCCGGGAGTACTTGTGTGTGTGCGGGCAGGAGAGG 364

Db 385 ATCTGTGGCTGGCTATGGCAGATTAACGGGGTGTACTCATGTGTGTGCGGAGGAGAGG 444

QY 365 ACCTCAGCAGCTCACCATC 385

Db 445 ACCTCAGCTACACTCACTGTC 465

RESULT 10

ABN21414

ID ABN21414 standard; cDNA; 1005 BP.

XX AC ABN21414;

XX 24-JUN-2002 (first entry)

XX Human ORFX polynucleotide sequence SEQ ID NO:11305.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;

KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;

KW degenerative disorder; osteoarthritis; neurodegenerative disorder;

KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;

KW hypertension; hypothyroidism; cholesterol ester storage disease;

KW immune deficiency; immune disorder; infectious disease;

KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;

KW myasthenia gravis; gene; ss.

XX Homo sapiens.

OS WO200192523-A2.

PN 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US010836.

PR 30-MAY-2000; 2000US-0206132P.

PR 29-AUG-2000; 2000US-0228716P.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach MD;

PI WPI: 2002-106308/14.

DR P-PSDB; ABP05662.

XX Novel human polypeptides and polynucleotides useful for diagnosing,

PT preventing and treating cardiovascular disease, neurodegenerative,

PT hyperproliferative disorders and autoimmune disorders.

XX Disclosure; SEQ ID NO 11305; 1037pp; English.

XX The present invention describes substantially purified human proteins

CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1

CC in the specification). ABN15762 to ABN27252 encode the human ORFX

CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for

CC treating or preventing a pathology associated with an ORFX-associated

CC disorder in humans, and in the manufacture of a medicament for treating a

CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide

CC sequences can be used in gene therapy. ORFX sequences can be used in the

CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,

CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,

CC osteoarthritis, neurodegenerative disorders, disorders related to organ

CC transplantation, cardiovascular diseases, diabetes mellitus, systemic

CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester

CC storage disease, various immune deficiencies and disorders, infectious

CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid

CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host

CC with a disease state; and (3) identifying polynucleotides and
CC polypeptides that predict compound sensitivity or resistance of cells
CC associated with a disease state, comprising subjecting the plurality of
CC cell lines to one or more compounds, analysing the expression pattern of
CC a microarray of polynucleotides or polypeptides, and selecting
CC polynucleotides or polypeptides that predict the sensitivity or
CC resistance of cells associated with a disease state by using the
CC expression pattern of the microarray. The polynucleotides and
CC polypeptides have cytostatic activities, and can be used in gene therapy.
CC The polynucleotides and polypeptides are useful in predicting the
CC activity of compounds that interact with protein tyrosine kinases and/or
CC protein tyrosine kinase pathways. These may be used in determining drug
CC sensitivity in patients to allow the development of individualized
CC genetic profiles which aid in treating diseases and disorders (e.g.
CC cancer) based on patient response at a molecular level. The present
CC sequence is used in the exemplification of the present invention.
XX
SQ Sequence 5382 BP; 943 A; 1699 C; 1798 G; 942 T; 0 U; 0 Other;

| Query Match | 12.4%; | Score 47.6; | DB 9; | Length 5382; |
|-----------------------|--|------------------|-----------|--------------|
| Best Local Similarity | 49.0%; | Pred. No. 0.038; | | |
| Matches 190; | Conservative 0; | Mismatches 189; | Indels 9; | Gaps 2; |
| 7 | CCTGAGCGCAGGACGGGGCTGTGTGCGAGCTGTCAGATCTGTGGCTGTGGCTGTGCAGATGC | 66 | | |
| Db | | | | |
| 2344 | CCTGCAGAAGAAACACCTGTCGCGCGCTGGTGTGTCGCCGTGTGCAGCTCGAGGACTC | 2403 | | |
| 67 | TGGGGAGTACTCTGTGTGTGTGTGGGAGGAGAGGACCTCTGCCACTCTCACCTGTGAAGGC | 126 | | |
| Db | | | | |
| 2404 | CGCGGAGTACTGTGTGTGAATTGACGATGATCGGCCCTCTCTCACTGTCACTGTACACGACAGA | 2463 | | |
| 127 | CTGTCCAGCGCAAGTTACACAGAGGTCT-----GAGGAATGAAGAGGCCCTGGAAGGEGC | 180 | | |
| Db | | | | |
| 2464 | ACCCGAGTCGGGATCATATACCTCGCATGAGTGACCTTGATGCCGTGACCTTGA | 2523 | | |
| 181 | CACAGCCATGTTGTGTGTGAACGTGA---GCAAGTGGCCCTGTGTGAGTGTGAGGAAAGG | 237 | | |
| Db | | | | |
| 2524 | GTGTGTGGTCTGATGTGTGAACCTGTCTCGGAGGATGCCCTGTGCGCTGTGTACAAAGGA | 2583 | | |
| 238 | GCCCGAGAACTCAGAGATGGGACACATACATCCTGAGCAGGAGGGGACCAAGTGTGA | 297 | | |
| Db | | | | |
| 2584 | TGGGCTGGAAGTGGAGGAGACGAGGCCCTGTGTCTGGAGAGGATGGGCCACGCTCCG | 2643 | | |
| 298 | GCTGTGAGATGTGGCCCTGGCCATGGCGGACGCCGGGAGTACTTGTGTGTGTGTCGGGCA | 357 | | |
| Db | | | | |
| 2644 | CCTGGTGTACCTGCTGCTCAGCCCGAGGACGGGGCGAGTTTGTATGTATGCTGTGAGA | 2703 | | |
| 358 | GGAGAGCACTCAGCCACCGTCCACATC | 385 | | |
| Db | | | | |
| 2704 | TGACTCGGCCCTTCTTCACTGTCACTGTC | 2731 | | |
| 358 | GGAGAGCACTCAGCCACCGTCCACATC | 385 | | |
| Db | | | | |
| 2704 | TGACTCGGCCCTTCTTCACTGTCACTGTC | 2731 | | |
| 358 | GGAGAGCACTCAGCCACCGTCCACATC | 385 | | |
| Db | | | | |
| 2704 | TGACTCGGCCCTTCTTCACTGTCACTGTC | 2731 | | |

| | |
|-----------|---|
| RESULT 14 | |
| AAS05390 | |
| ID | AAS05390 standard; DNA; 81940 BP. |
| XX | |
| XX | |
| AC | AAS05390; |
| XX | |
| XX | |
| DT | 24-OCT-2001 (first entry) |
| XX | |
| XX | |
| DE | Human titin (connectin) gene sequence. |
| XX | |
| XX | |
| KW | Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B; |
| KW | titin-related disease; zebrafish; heart failure; heart disease; ds. |
| XX | |
| XX | |
| CS | Homo sapiens. |
| XX | |
| XX | |
| Key | Location/Qualifiers |
| FT | 133..80913 |
| CDS | /*tag= a |
| FT | /product= "titin" |
| FT | |
| XX | |
| XX | |
| FN | WO200151666-A1. |

19-JUL-2001.
12-JAN-2001; 2001WO-US001212.
12-JAN-2000; 2000US-0175787P.
(GEO) GEN HOSPITAL CORP.
Fishman MC;
WPI; 2001-451869/48.
P-PSDB; AAU05396.
Determining if a subject has or is at risk of developing a titin-related disease or condition, particularly heart failures, comprises detecting the presence of a mutation in the titin gene.
Disclosure; Page 35-57; 114pp; English.
The present sequence encoding for human titin (also known as connectin) is described in an invention relating to a novel method for determining whether a subject has or is at risk of developing a titin-related disease or condition. The method comprises analysing a nucleic acid of a sample from the subject and detecting the presence of a mutation (e.g. the pickwick mutation in the cardiac specific exon N2B) in the titin gene, which indicates that the subject has or is at risk of developing a titin-related disease. The zebrafish which has a phenotype similar to mammalian heart failure is used as a model. The method is useful for detecting an increased likelihood of heart disease, such as heart failure, in a patient, so that appropriate intervention can be instituted before any symptoms occur. The method may also be used to facilitate determination of etiology of an existing heart condition, such as heart failure, to identify compounds that can be used to treat or prevent heart conditions, in prenatal genetic screening, e.g. to identify parents who may be carriers of a recessive titin mutation. Compounds identified using the methods may be used to treat patients that have or are at risk of developing heart disease, e.g. heart failure

| Query Match | 12.2% | Score 46.8 | DB 4 | Length 81940 | |
|-----------------------|--------------|--|----------------|--------------|--------|
| Best Local Similarity | 48.8% | Fred. No. 0.1 | | | |
| Matches 157 | Conservative | 0 | Mismatches 162 | Indels 3 | Gaps 1 |
| QY | 66 | CTGGGGAGTACTCCCTGTGTGTGGGAGGAGAGGAGCTCTGCCACTCTCACCGTGAAGG | 125 | | |
| Db | 16091 | CAGGAGAGGTTTCCTTCAGGCTGCTAATGCCAATCTGCAGCCAACTCGAAGTGAAG | 16150 | | |
| QY | 126 | CCCTGCCAGCCAAAGTTCACAGAGGTTCTGAGGAATGAAGAGGCCGTGGAAAGGGGCCACAG | 185 | | |
| Db | 16151 | AATTGCCCTCTTATCTTCATCACACTCTCAGTGATGTTAAAGTCTTCGAGAAAGATGAGG | 16210 | | |
| QY | 186 | CCATGTTGTGGTGTAATGACAGCAAGTGGCC---CCTGTGGAGTGGAGGAAGGGGCCCG | 242 | | |
| Db | 16211 | CTAAGTTTGAGTGTGAAGTATCCAGGGAGGCCAAAACATTCCTGTGCTTAAAGGAACC | 16270 | | |
| QY | 243 | AGAACTCAGAGATGGGACAGATACATCTCTGAGCGAGGGAGGCCACGAGTGTGAGCTGC | 302 | | |
| Db | 16271 | AGGAATCACAGGTGATCACAGATTGAGCTTATAAGGATGGCACTAAGCATTCAATGG | 16330 | | |
| QY | 303 | AGATCTGTGGCCTGGCCATGGCGGACGCCGGGGAGTACTTGTGTGTGTGGCGGACAGAGA | 362 | | |
| Db | 16331 | TGATCAAGTCAGCTGCTTTGAAGATGAAGCAAAATACATGTTTGAAGCTGAAGATAAGC | 16390 | | |
| QY | 363 | GGACCTCAGCCACGCTCACCAT | 384 | | |
| Db | 16391 | ACACAAGTGGCAAACTGATCAT | 16412 | | |

RESULT 15
ABK64829
ID ABK64829 standard; DNA; 81940 BP.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 12:53:28 ; Search time 441.18 Seconds
(without alignments)
10303.209 Million cell updates/sec

Title: US-10-077-130-4_COPY_1_1070

Perfect score: 1070

Sequence: 1 tgcctaccagcagccacac.....tagtgcgagccgcgggt 1070

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002s:*
7: Geneseq2003as:*
8: Geneseq2003bs:*
9: Geneseq2003cs:*
10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 1070 | 100.0 | 24120 | 7 ABX11642 | Abx11642 Human ser |
| 2 | 255 | 23.8 | 2884 | 7 ADA53433 | Ada53433 Human cod |
| 3 | 197.6 | 18.5 | 9591 | 4 AAS28906 | Aas28906 Human imm |
| 4 | 197.6 | 18.5 | 9591 | 4 AAL03355 | Aal03355 Human rep |
| 5 | 197.6 | 18.5 | 9591 | 9 ADB31747 | Adb31747 Human nov |
| 6 | 197.6 | 18.5 | 12415 | 4 AAS28907 | Aas28907 Human imm |
| 7 | 197.6 | 18.5 | 12415 | 4 AAL03356 | Aal03356 Human rep |
| 8 | 197.6 | 18.5 | 12415 | 9 ADB31748 | Adb31748 Human nov |
| 9 | 197.6 | 18.5 | 20565 | 4 AAK89083 | Aak89083 Human dig |
| 10 | 197.6 | 18.5 | 20565 | 4 AAS28908 | Aas28908 Human imm |
| 11 | 197.6 | 18.5 | 20565 | 4 AAL03357 | Aal03357 Human rep |
| 12 | 197.6 | 18.5 | 20565 | 9 ADB31749 | Adb31749 Human nov |
| 13 | 126 | 11.8 | 5382 | 9 ADD14722 | Add14722 Human src |
| 14 | 117.2 | 11.0 | 471 | 2 AAV88238 | Aav88238 EST clone |
| 15 | 99.6 | 9.3 | 9975 | 7 AAL61173 | Aal61173 Actinosyn |
| 16 | 99.6 | 9.3 | 82746 | 7 AAL61173 | Aal61173 Actinosyn |
| 17 | 90.2 | 8.4 | 114955 | 2 AAX53491 | Aax53491 Human ade |
| 18 | 86.6 | 8.1 | 609 | 4 ABA06472 | Abao6472 Human cdn |
| 19 | 86.6 | 8.1 | 609 | 6 ABA06472 | Abao6472 Human cdn |
| 20 | 83.2 | 7.8 | 114955 | 2 AAX53491 | Aax53491 Human ade |
| 21 | 78.6 | 7.3 | 14055 | 7 AAL61170 | Aal61170 Actinosyn |
| 22 | 74.8 | 7.0 | 3015 | 3 ARA29367 | Ara29367 Human zsi |
| 23 | 73.4 | 6.9 | 65140 | 4 AAD17184 | Aad17184 Streptomy |

| | | | | | | |
|----|------|-----|--------|---|-----------|---------------------|
| 24 | 73.4 | 6.9 | 125401 | 4 | AAD17186 | Aad17186 Streptomy |
| 25 | 72.8 | 6.8 | 3183 | 7 | ABZ77314 | Abz77314 Nucleotid |
| 26 | 70.5 | 6.6 | 4020 | 2 | AA791361 | Aat91361 Orf virus |
| 27 | 70.2 | 6.6 | 1281 | 7 | ABZ66730 | Abz66730 Orthosomy |
| 28 | 70.2 | 6.6 | 15240 | 7 | ABZ66809 | Abz66809 Orthosomy |
| 29 | 70.2 | 6.6 | 109519 | 5 | AAS08693 | Aas08693 Micromono |
| 30 | 70 | 6.5 | 4341 | 7 | ABT32151 | Abt32151 Benzodiaz |
| 31 | 70 | 6.5 | 32539 | 7 | ABT32129 | Abt32129 Benzodiaz |
| 32 | 68.6 | 6.4 | 1614 | 5 | AAH26499 | Aah26499 Human low |
| 33 | 68.6 | 6.4 | 5760 | 6 | ABST78661 | Abst78661 M. echino |
| 34 | 68.6 | 6.4 | 12425 | 5 | AAH26495 | Aah26495 Human low |
| 35 | 68 | 6.4 | 1226 | 4 | AAH07456 | Aah07456 Human rep |
| 36 | 68 | 6.4 | 1776 | 9 | ADE03416 | Ade03416 Human imm |
| 37 | 68 | 6.4 | 1779 | 6 | AAD16344 | Aad16344 Human sbg |
| 38 | 68 | 6.4 | 1779 | 6 | ABQ86161 | Abq86161 Novel hum |
| 39 | 68 | 6.4 | 1811 | 6 | ABK48387 | Abk48387 DNA encod |
| 40 | 68 | 6.4 | 1860 | 6 | ABK92037 | Abk92037 DNA encod |
| 41 | 68 | 6.4 | 2026 | 9 | ADB62582 | Adb62582 Human cdn |
| 42 | 68 | 6.4 | 2080 | 6 | ABK62080 | Abk62080 Human cdn |
| 43 | 68 | 6.4 | 2222 | 6 | ABK62081 | Abk62081 Human cdn |
| 44 | 68 | 6.4 | 2300 | 6 | AAD41275 | Aad41275 Human LP |
| 45 | 68 | 6.4 | 2338 | 7 | ABQ77432 | Abq77432 Human CGD |

ALIGNMENTS

RESULT 1

ABX11642
ID ABX11642 standard; cDNA; 24120 BP.

XX
AC ABX11642;

XX
DT 09-MAY-2003 (first entry)

XX
DE Human serine/threonine or protein kinase 12599, cDNA.

XX
KW Human; ss; gene; serine/threonine kinase; protein kinase; 12599;
KW cardiovascular disease; heart failure; myocardial infarction;
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;
KW hemolytic anaemia; thrombocytopenia; leukaemia; Hodgkin's disease;
KW protein kinase disorder; cellular proliferative disorder; cancer;
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;
KW multiple sclerosis.

XX
OS Homo sapiens.

XX
FH Key Location/Qualifiers

FT 5'UTR 1..71

FT CDS /tag= a

FT /tag= b

FT /product= "Kinase 12599"

FT /note= "This CDS is specifically claimed in claim 2"

3'UTR 23979..24120

FT /tag= c

XX
US2002168742-A1.

XX
14-NOV-2002.

XX
15-FEB-2002; 2002US-00077130.

XX
15-FEB-2001; 2001US-0269201P.

XX
(MILL-) MILLENNIUM PHARM INC.

XX
Kapeller-Libermann R, Acton SL;

XX
WPI; 2003-298729/29.

XX
P-PSDB; ABG76187.

PT Novel isolated human protein kinase, designated 59079 or 12599
PT polypeptide, useful as diagnostic and therapeutic agents for preventing
PT cardiovascular diseases, proliferative disorders, and protein kinase
PT disorders.

PS Claim 2; Page 58-84; 119pp; English.

XX The invention relates to an isolated human serine/threonine or protein
XX kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule
XX comprising at least 85% identity to the nucleic acids appearing as
XX ABX11641 and ABX11642 or their complement, a naturally occurring variant
XX of the kinases or their fragments. Also included are a non-human host
XX cell containing the nucleic acids, an antibody specific for the proteins,
XX identifying a compound which binds to the kinase (by contacting the
XX kinase or a cell expressing the kinase with a test compound and
XX determining whether the kinase binds to the test compound) and modulating
XX the activity of kinase using the identified compound. The kinases and
XX their encoding nucleic acids are useful as diagnostic and therapeutic
XX agents for preventing a disease or condition associated with an aberrant
XX or unwanted 59079 or 12599 activity in a subject, including
XX cardiovascular diseases such as heart failure, and myocardial infarction;
XX disorders involving blood vessels such as atherosclerosis, and Kaposi's
XX sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia,
XX Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders
XX such as cancer; and protein kinase disorders such as autoimmune
XX disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,
XX rheumatoid arthritis, and multiple sclerosis (many examples of diseases
XX and disorders are included in the specification). The kinases, their
XX encoding nucleic acids and antibodies are useful in screening assays,
XX detection assays (e.g. forensic biology), and predictive medicine (e.g.
XX diagnostic assays, prognostic assays, and monitoring clinical trials and
XX pharmacogenomics). The kinases and their encoding nucleic acids are
XX useful as query sequences to perform a search against public databases to
XX identify other family members or related sequences. The present sequence
XX encodes the kinase 12599

SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;

Query Match 100.0%; Score 1070; DB 7; Length 24120;
Best Local Similarity 100.0%; Pred. No. 6.8e-162;
Matches 1070; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCCTACAGCAGCCACACTCCCGCGCTGCCAGAGCCCCCATAGAGAGGTCCCCG 60
DB 1 TGCCTACAGCAGCCACACTCCCGCGCTGCCAGAGCCCCCATAGAGAGGTCCCCG 60
QY 61 CCGCCACCGTATGATCAGGCACAGTTCAGCGGGGCGCGCGCTTCTCACCGGCCCA 120
DB 61 CCGCCACCGTATGATCAGGCACAGTTCAGCGGGGCGCGCGCTTCTCACCGGCCCA 120
QY 121 AGGCTTCTGCTGCTGGGCAAGGACGACCCCTCAGTGCAGATCGTGGTAATC 180
DB 121 AGGCTTCTGCTGCTGGGCAAGGACGACCCCTCAGTGCAGATCGTGGTAATC 180
QY 181 CCAGCCACAGGTGAGTGGGAGAGGACGACGAGCGGTGACGCGCGCGCGCTTCC 240
DB 181 CCAGCCACAGGTGAGTGGGAGAGGACGACGAGCGGTGACGCGCGCGCGCTTCC 240
QY 241 GTCTGGCCACGAGCGGCGACCTTACCGCTTACTATCTGGACCTGGCGCTGGCGACA 300
DB 241 GTCTGGCCACGAGCGGCGACCTTACCGCTTACTATCTGGACCTGGCGCTGGCGACA 300
QY 301 GTGGGCAATACGTGTGCGCGCGGCAATGCAATGCGAGGCGCTTCTGCTGGCGGCC 360
DB 301 GTGGGCAATACGTGTGCGCGCGGCAATGCAATGCGAGGCGCTTCTGCTGGCGGCC 360
QY 361 TGCAGGTGAGCGCGAGCGCGGTGCGCGAGGAGCGCGCACTTCTGCTGGCGGCCA 420
DB 361 TGCAGGTGAGCGCGAGCGCGGTGCGCGAGGAGCGCGCACTTCTGCTGGCGGCCA 420
QY 421 CGTCCATCCGCGTGGCGAGGCGCTCAGAGGCCACTTTCGCTGCGGTGGGTCC 480
DB 421 CGTCCATCCGCGTGGCGAGGCGCTCAGAGGCCACTTTCGCTGCGGTGGGTCC 480

QY 481 CGAGCCCGCAGTGTGCTGCTCAAGAGCGCGCGCTGGGTGAGCCCGAGCGGCCCC 540
DB 481 CGAGCCCGCAGTGTGCTGCTCAAGAGCGCGCGCTGGGTGAGCCCGAGCGGCCCC 540
QY 541 CGCTCGCGTGTGAGAGCTCGCGAGGCAAGTGTGCTGCGCATTCGGCGCGCGCGCC 600
DB 541 CGCTCGCGTGTGAGAGCTCGCGAGGCAAGTGTGCTGCGCATTCGGCGCGCGCGCC 600
QY 601 CGCAGCGCGCACCTTACGAGTCTCGCGCGAGAACCCGCTGGGCGCTGCCAGCGCCG 660
DB 601 CGCAGCGCGCACCTTACGAGTCTCGCGCGAGAACCCGCTGGGCGCTGCCAGCGCCG 660
QY 661 CGGCGCTAGTGTGAGTCTCGAGCGCGGAGACAGGCGCGCGCGCGCGCGCTCCACGG 720
DB 661 CGGCGCTAGTGTGAGTCTCGAGCGCGGAGACAGGCGCGCGCGCGCGCGCTCCACGG 720
QY 721 CGGCGCTCTGCGCGCACCTGCGAGCGCGCGCGAGGCTATGCGCGCGCGCGCGCC 780
DB 721 CGGCGCTCTGCGCGCACCTGCGAGCGCGCGCGAGGCTATGCGCGCGCGCGCC 780
QY 781 CTTACCCCGCAGCACCGGACCGCGCGCTGCTGAGAGGAGCGCGCGCGCGCGCC 840
DB 781 CTTACCCCGCAGCACCGGACCGCGCGCTGCTGAGAGGAGCGCGCGCGCGCGCC 840
QY 841 TCAGTCTGTGCTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
DB 841 TCAGTCTGTGCTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
QY 901 TGGTACCGAGCG 960
DB 901 TGGTACCGAGCG 960
QY 961 AGATCTCTTCTGCAAGCAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
DB 961 AGATCTCTTCTGCAAGCAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
QY 1021 TGGGCGAGACCTTACAGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1070
DB 1021 TGGGCGAGACCTTACAGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1070

RESULT 2

ADA53433/c
ID ADA53433 standard; cDNA; 2884 BP.
XX
AC ADA53433;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 1001.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

| | | |
|----|------------------------------|------------------|
| PR | 20-OCT-2000; | 2000US-0240960P. |
| PR | 20-OCT-2000; | 2000US-0241221P. |
| PR | 20-OCT-2000; | 2000US-0241785P. |
| PR | 20-OCT-2000; | 2000US-0241786P. |
| PR | 20-OCT-2000; | 2000US-0241787P. |
| PR | 20-OCT-2000; | 2000US-0241808P. |
| PR | 20-OCT-2000; | 2000US-0241809P. |
| PR | 20-OCT-2000; | 2000US-0241826P. |
| PR | 01-NOV-2000; | 2000US-0244617P. |
| PR | 08-NOV-2000; | 2000US-0246474P. |
| PR | 08-NOV-2000; | 2000US-0246475P. |
| PR | 08-NOV-2000; | 2000US-0246476P. |
| PR | 08-NOV-2000; | 2000US-0246477P. |
| PR | 08-NOV-2000; | 2000US-0246478P. |
| PR | 08-NOV-2000; | 2000US-0246523P. |
| PR | 08-NOV-2000; | 2000US-0246524P. |
| PR | 08-NOV-2000; | 2000US-0246525P. |
| PR | 08-NOV-2000; | 2000US-0246526P. |
| PR | 08-NOV-2000; | 2000US-0246527P. |
| PR | 08-NOV-2000; | 2000US-0246528P. |
| PR | 08-NOV-2000; | 2000US-0246532P. |
| PR | 08-NOV-2000; | 2000US-0246609P. |
| PR | 08-NOV-2000; | 2000US-0246610P. |
| PR | 08-NOV-2000; | 2000US-0246611P. |
| PR | 08-NOV-2000; | 2000US-0246613P. |
| PR | 17-NOV-2000; | 2000US-0249207P. |
| PR | 17-NOV-2000; | 2000US-0249208P. |
| PR | 17-NOV-2000; | 2000US-0249210P. |
| PR | 17-NOV-2000; | 2000US-0249211P. |
| PR | 17-NOV-2000; | 2000US-0249212P. |
| PR | 17-NOV-2000; | 2000US-0249213P. |
| PR | 17-NOV-2000; | 2000US-0249214P. |
| PR | 17-NOV-2000; | 2000US-0249215P. |
| PR | 17-NOV-2000; | 2000US-0249216P. |
| PR | 17-NOV-2000; | 2000US-0249217P. |
| PR | 17-NOV-2000; | 2000US-0249218P. |
| PR | 17-NOV-2000; | 2000US-0249244P. |
| PR | 17-NOV-2000; | 2000US-0249245P. |
| PR | 17-NOV-2000; | 2000US-0249264P. |
| PR | 17-NOV-2000; | 2000US-0249265P. |
| PR | 17-NOV-2000; | 2000US-0249292P. |
| PR | 17-NOV-2000; | 2000US-0249299P. |
| PR | 17-NOV-2000; | 2000US-0249300P. |
| PR | 01-DEC-2000; | 2000US-0250160P. |
| PR | 01-DEC-2000; | 2000US-0250391P. |
| PR | 05-DEC-2000; | 2000US-0251030P. |
| PR | 05-DEC-2000; | 2000US-0251388P. |
| PR | 05-DEC-2000; | 2000US-0256719P. |
| PR | 06-DEC-2000; | 2000US-0251479P. |
| PR | 08-DEC-2000; | 2000US-0251856P. |
| PR | 08-DEC-2000; | 2000US-0251868P. |
| PR | 08-DEC-2000; | 2000US-0251869P. |
| PR | 08-DEC-2000; | 2000US-0251989P. |
| PR | 08-DEC-2000; | 2000US-0251990P. |
| PR | 11-DEC-2000; | 2000US-0254907P. |
| PR | 01-JAN-2001; | 2000US-0259567P. |
| XX | | |
| PA | (HUMA-) HUMAN GENOME SI INC. | |

CC the presence or absence of a mutation in a DNA sequence or determining
CC the presence or amount of expression of the protein. Alternatively the
CC identification of a binding partner to a sequence allows determination of
CC changes in protein activity. The sequences can be used as research tools
CC for receptors or other signal transduction pathway proteins that interact
CC with the polypeptides of the invention and can be used to treat, prevent
CC or diagnose various types of disorders such as neurological disorders,
CC cardiovascular disorders, gastrointestinal disorders, reproductive
CC disorders, immune system disorders, renal disorders, muscular disorders,
CC pulmonary disorders, proliferative disorders and cancer. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)
xx
SQ Sequence 9591 BP; 2036 A; 2759 C; 2671 G; 2135 T; 0 U; 0 Other;

Isolated novel immunoglobulin polypeptide for monitoring the presence and progression of diseases and for diagnosis.

Claim 1; SEQ ID NO 268; 551pp; English.

Sequences AAS28878-AAS28926 represent genomic DNA molecules which encode the immunoglobulin polypeptides of the invention. The polynucleotides and polypeptides can be used to diagnose a pathological condition or a susceptibility to a pathological condition in a subject by determining

Db 782 CCGCGCGGTGGTGGAGCCGCTCAAGTGCAGCGCTTAAGACCTTCTGGGTGAACGAGGGCA 841
Qy 829 AGCAGCGCGGCTCAGCTGTGTACGTACCGCGCGAGCCCAAGCCGAGACCGGTGTGGAAGA 888
Db 842 AGCAGCGCAAGTTCGCTGTACGTGATGGCAAGCCGAGCCGAGATCGAATGGCACT 901
Qy 889 AGGACGGCGAGCTGGTACCGAGGGCGCGGCCACAGTGGTGTAGGAGGAGCGCGCAGAGA 948
Db 902 GGGAGGGCGCGCGCGCTCTCCCGACCGCGCGCGCTCATGTACCGGACCGCGAGCGCGG 961
Qy 949 ACTTCGTGCTCAAGATCTCTCTGCAAGCAGTCCGACCGCGGCTCTACACCTGCAACGG 1008
Db 962 GCTTCGTGCTCAAGTCTTACTGCGAGCCAGGATCGTGGCTCTAGCTGTGGCGG 1021
Qy 1009 GGTCCAACTGTGGGCGACACCTACAGCTCTGTGCTGTGAGTGGCGGAGCCCGCGG 1068
Db 1022 CGCGAACTCGCGGGCGAGCGCTCAGTGCCTGCGAGCTGCAGTGAAGGTACGCGCG 1081

RESULT 4

AAL03355

ID AAL03355 standard; DNA; 9591 BP.

XX AC

XX AAL03355;

XX DT 21-NOV-2001. (first entry)

XX DE Human reproductive system related antigen DNA SEQ ID NO: 6043.

XX KW Human; reproductive system related antigen; reproductive system disorder;
XX KW cancer; gene therapy; ds.
XX OS Homo sapiens.

XX PN WO20015320-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001339.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 28-JUN-2000; 2000US-0214886P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

XX PR 07-JUL-2000; 2000US-0216880P.

XX PR 11-JUL-2000; 2000US-0217487P.

XX PR 11-JUL-2000; 2000US-0217496P.

XX PR 14-JUL-2000; 2000US-0218290P.

XX PR 26-JUL-2000; 2000US-0220963P.

XX PR 26-JUL-2000; 2000US-0220964P.

XX PR 14-AUG-2000; 2000US-0224518P.

XX PR 14-AUG-2000; 2000US-0224519P.

XX PR 14-AUG-2000; 2000US-0225213P.

XX PR 14-AUG-2000; 2000US-0225214P.

XX PR 14-AUG-2000; 2000US-0225266P.

XX PR 14-AUG-2000; 2000US-0225267P.

XX PR 14-AUG-2000; 2000US-0225268P.

XX PR 14-AUG-2000; 2000US-0225270P.

XX PR 14-AUG-2000; 2000US-0225447P.

XX PR 14-AUG-2000; 2000US-0225577P.

XX PR 14-AUG-2000; 2000US-0225756P.

XX PR 14-AUG-2000; 2000US-0225759P.

XX PR 18-AUG-2000; 2000US-0226279P.

XX PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.

PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 05-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254087P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-465570/50.
 XX
 XX Isolated nucleic acid molecule encoding a reproductive system antigen is
 PT used in preventing, treating or ameliorating a medical condition.
 TT
 PS Disclosure; SEQ ID NO 6043; 1297pp + Sequence Listing; English.
 XX
 XX The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a genomic sequence encoding a
 CC protein of the invention
 XX
 SQ Sequence 9591 BP; 2026 A; 2759 C; 2671 G; 2135 T; 0 U; 0 Other;
 Query Match 18.5%; Score 197.6; DB 4; Length 9591;
 Best Local Similarity 51.7%; Pred. No. 4.8e-23;
 Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;
 YQ 7 CCACGAGCCACACTCCGGCGCTGCCCCAGAGCCCCCATAGAGAGAGTCCCCCGCCGCA 66
 DB 2 CCTCCGACACCCCTTGGCCCCCACCCTTGGCGCTGCAGGCGGTGGCGCGCCGGA 61
 YQ 67 CCGTCATGATACGCCACAGTTACGCGGGCGCCCGCTTTCTACCCCGGCCCAAGCCT 126
 DB 62 TGAAGGCGAGTCCGGGGGATCAGGGAGCCCCCGCTTCTTCCGCTTCCCGCGCGCTG 121
 YQ 127 TCGTGGTGTGGTGGGCAAGACCCACCTCAGTCCAGATCGTGGTAATCCACGC 186
 DB 122 TCGGGTGGTAAGTGGCGCCGAGCCGAGTCAAGTCCGTGCTGGGGAGCCCGCC 181
 YQ 187 CACAGGTGAGTGGAGAGGACACAGCAGCGGTGACGCGCGCGCGCTCCGCTCG 246
 DB 182 CTGTAGTGGTGGAGAGGGGGGAGCAGCTGGCGGCTCGGAACGCTGAGCTTC 241
 YQ 247 CCCAGGACGGCCTTACCGCTCACTCTGACCTGGCGCTGGGGCGACAGTGGC 306

DB 242 CGGCGACGCGCGGACGACCGGCTGTGTGACCGCGCACTGCCACACGACGCGGGG 301
 QY 307 AATAGTGTGCGCGCGCGGCAATGCCATAGCGAGGGCTTGGCTGGCTGGGCTGACGG 366
 DB 302 TCTAGTGTGCGCGCGCGCAACGCGCGCGGAGGCTACGCGCGCGCGCGCTACCG 361
 QY 367 TGGACGCGGAGCGCGCTGCGCGCGGAGCGCGCGCTTCTGCTGCGGCGCCACGTCGA 426
 DB 362 TGTGGAGCGCGCGCTTCCGACCCCGAGCTGCAGCCCGCGCGCGCTGCTGCTGCG 421
 QY 427 TCCGGTGGCGAGGCTCAGAGGCACTTCCGCTGCGG-----GTTGGTGGCTCC 480
 DB 422 CGGGTCCGGGAGGCGCGCGCTTCTTCTCAGCGGCGCTCGATCCAGTGGGTGCTGC 481
 QY 481 CGAGCGCGGAGTGAAGTGGTCCAGGACGCGCGCGCTTGGGTGAGCGCGCGCGCG 540
 DB 482 GGGGGCGAGTGTGTGCTGAGCTGCGCGGCGGGGCTTCCCGAGCGCCACACTGTACT 541
 QY 541 GGTGCGCTGAGAGCTCGGCGAGGAAAGTGTGCTGCGGCTTCCGCGCGCGCGCGCG 600
 DB 542 GGGAGAGGACGGATGCCCTGACGAAAGTGTGGACAGCAGCACTTCCGCTCCAGC 601
 QY 601 GCGAGCGCGCACTTACAGGTCCGCGCGGAGAGCCGCTGGG--GCTGCCAGCGCGC 658
 DB 602 CGGCGCGCGGAGACGCGCGCGGCGAGCTTGGCACTTGGCATCTTGGCGCTCGC 661
 QY 659 CGCGCGCTAGTG-GTGGACTCGGACGCGCGGACAGCGCGCGCGCGCGCGCGCTCCA 717
 DB 662 TCCGGATTCCGGCGTCTTACGTGCGCACGCGCGCAACGCGCACGCGCGCGCG 721
 QY 718 CGCGCGCTCTTGGCGCACTTGCAGCGCGCGCGGAGGCTATCGCGCGCGGAGCGCG 777
 DB 722 GGGCGCTGCTCAGGTGCACAGCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCG 781
 QY 778 CCGCTC-----ACCGCGCAGCAGCGCGCGCGCGCTGACCGTGTGCTGAAAGCGCA 828
 DB 782 CCGCGCGGTGTGGAGCGCGCTCAAGTGGCGCTTAAGACTTCTGGTGAACGAGGCA 841
 QY 839 AGCAGCGCGCTCAGTGTGCTGACCGCGGAGCGCGCGCGCGCGCGCGCGTGTGAAGA 886
 DB 842 AGCAGCGCAAGTCCGCTGTACGTGATGGCAAGCGCGCGCGCGCGCGCGTGTGCACT 901
 QY 889 AGGAGCGCGCACTGTGTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGA 948
 DB 902 GGGAGCGCGCGCGCTGCTCCGAGCGCGCGCGCGCGCTCATGTACCGGACCGCGCG 961
 QY 949 ACTTGTGCTCAAGATCCTTCTTCAAGAGTTCGACCGCGCGCGCGCTTACCTGTGACGG 1008
 DB 962 GTTCTGTCTCAAGTGTCTTACTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1021
 QY 1009 CGTCCACCTCGTGGCGCAGACTACAGCTGTGCTGTGCTGAGTGGCGGAGCGCGCG 1068
 DB 1022 CGCGCACTCGCGCGCGCGCGCTCAGTCCGCTGACGCTGCACGTAAGAGTACGCGCG 1081
 RESULT 5
 ADB31747
 ID ADB31747 standard; DNA; 9591 BP.
 XX
 AC ADB31747;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Human novel protein DNA SEQ ID NO 268.
 XX
 DE gene therapy; human; immunoglobulin; cancer; lung cancer; leukaemia;
 KW ovarian cancer; epithelial cancer; hyperplasia; Gaucher's disease; AIDS;
 KW arhythmia; cardiac oedema; ischaemia; pneumonia; cystic fibrosis;
 KW asthma; sarcoidosis; rhinitis; anaemia; inflammation; sinusitis;
 KW chronic obstructive pulmonary disease; infectious disease; gene; ds.
 XX
 OS Homo sapiens.
 XX

FN US2003077606-A1.
XX 24-APR-2003.
XX
XX
XX 07-MAR-2002; 2002US-00091438.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216860P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226888P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227709P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 06-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764879.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX

PI Rosen CA, Ruben SM, Barash SC;
XX WPI; 2003-615993/58.
XX
XX New human immunoglobulin superfamily of polypeptide and genes, useful for
XX treating, preventing or diagnosing e.g. cancers (lung cancer, leukemia,
XX ovarian epithelial cancer, etc.), hyperplasia, ischemia, pneumonia or
XX AIDS.
XX
XX Disclosure; SEQ ID NO 268; 213pp; English.
XX
XX The invention relates to an isolated polypeptide, which comprises the
XX human immunoglobulin superfamily of proteins. The polypeptide or
XX polynucleotide is useful for treating, preventing or ameliorating a
XX medical condition e.g. cancers (lung cancer, leukemia, ovarian
XX epithelial cancer, etc.), hyperplasia, Gaucher's disease, AIDS,
XX arhythmia, cardiac oedema, ischemia, pneumonia, cystic fibrosis,
XX asthma, sarcoidosis, rhinitis, anaemia, leukaemia, inflammatory
XX sinusitis, chronic obstructive pulmonary disease, infectious diseases
XX etc. The polypeptide or polynucleotide is also useful for diagnosing any
XX of these diseases or a susceptibility to the disease. The present
XX sequence represent a novel human protein DNA. Note: The sequence data for
XX this patent did not form part of the printed specification but was
XX obtained in electronic format direct from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=20030077606
XX
XX Sequence 9591 BP; 2026 A; 2759 C; 2671 G; 2135 T; 0 U; 0 Other;
XX
XX Query Match 18.5%; Score 197.6; DB 9; Length 9591;
XX Best Local Similarity 51.7%; Pred. No. 4.8e-23;
XX Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCACAGGCCACATCCGGCGGTGCCCCAGAGCCCCCATAGAGAGAGTCCCGCGGCCA 66
DB 2 CCCTCCGACACCCCTTGCCTCCCGCCACCGTTCCCGCTGCAGCGGTGCGCGCGCGCA 61
QY 67 CCGTCATGGATCAGCCACAGTTACAGCGGGCGCGCCCTTTCTCACCGCGCCCAAGCCT 126
DB 62 TGAAGGCGAGTCTGGGGGATCAGGGAGCCCGCGTCTCTCGGTTCGCGGCGCTG 121
QY 127 TCGTGTGTCGTGGGCAAGACCCACCTCAGTCTCCAGATCGTGGGTAATCCACGC 186
DB 122 TCGGGTGGTAAGTGGCGCGAGCCGAGCTCAAGTGCCTGCTGGGGAGCGCCGC 181
QY 187 CACAGGTGAGTGGGAGAGGACACAGCGCGGTGACGGCGCGCGCTCCGCTGG 246
DB 182 CTGTAGTGGTGGGAGAGGGCGGAGAGTGGCGGCTCGAACCGCTGAGCTTCC 241
QY 247 CCCAGGACGGACCTCTACCGCTCTACTCTCTGGACCTGGCGCTGGCGACAGTGGGC 306
DB 242 CGGCGGACGGCGGAGCAGCGCTGTGTGACCGCGCGCACTGCGCCACCGCGGGG 301
QY 307 AATACGTGTCCGCGCGCAATGCCATAGCGAGGCTTCGCTGCGTGGGCTCGAG 366
DB 302 TCTACGTGTCCGCGCGCCCAACGGCGCGGAGGCTTACGGCGGCGCGCGCTCACCG 361
QY 367 TGGACGGGAGGCGCGTGGCGCGAGCAGCGCGCGCACTTCTCTGCTGGCGCCACGTCCA 426
DB 362 TGCTGGAGCGCGCGCTCCGACCCGAGTGCAGCGCGCGCGCGCTGCCATCGC 421
QY 427 TCCGCGTGGCGAGGCTCAGAGCCACTTCCGCTGCGG-----CGTGGGTGGCTCC 480
DB 422 CGGGGTCCGGGAGGGCGCGCGCGCTTCTCTCAAGGGGCTCGATCCCATGGGTCTGC 481
QY 481 CGAGCGCGGAGTGAAGTGTGCTGCAAGGACGGCGCGCGCGCTGGGTGAGCGCGCGCC 540
DB 482 GGGGGGCGGAGTGTGCTGACGTGCGCGCGCGGGGGCTTCCCGAGCGCCACACTGACT 541
QY 541 GCGTGGCGGTGGAGAGCTCGCGGAGCGCAAGTGGCTGCGCATTCGGGCGCGCGCGCG 600
DB 542 GGGAGAGGAGCGGATCGCGCTTGGACGAAAGTGTGGGACAGCAGCCACTTGGCGCTCAG 601
QY 601 GCGACGGCGGACATTACAGGTTCGCGCGCGGAGAACCGCGTGGGC---GCTGCCAGCGCG 658

DB 602 CGGGCGCGCGAGGAGCGGCCCGCGGAGCTTGGCACTCGCATCTCGCGGCTCGGC 661
QY 659 CGCGGCGCTAGTG-GTGACTCGGAGCGCGGAGCACGGCGCGCGCGCGGACCTCCA 717
DB 662 TCCCGGATTCGGCGCTTACGTGTGCACGCCCGCAACGCCACCGCCACCGCAGGCG 721
QY 718 CGGGCGCGCTTCTGTGCGCACCTGTGAGCGCGCGCGCGAGGCTTATGCGCGCGCGCGCC 777
DB 722 GGGCGCTGCTCCAGGTGCACCCAGCCCCCGGAGAGCGCGCGCGCGCGCGCGCG 781
QY 778 CGGCTC-----ACGGCCAGCACCGSCACCGCACCTGACGGTGAATGAGGCA 828
DB 782 CGCGCCCGGTGGTGGAGCGCTCAAGTGCGCCCTTAAGACCTTGGGTGAACGAGGCA 841
QY 829 AGCAGCGCGGCTCAGCTGTACGTGACCGCGGAGCCCAAGCCGAGCGGTGTGGAAGA 888
DB 842 AGCAGCGCAAGTTCGCTGTCTACGTGTGCGCAAGCGCGCGAGATCGAATGGCACT 901
QY 889 AGGAGCGCGAGTGTGTACCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAGA 948
DB 902 GGGAGGCGCGCGCGCTGTCTCCGCGACCGCGCGCGCTCATGTACCGCGACCGCGCG 961
QY 949 ACTTGTGCTCAAGATCCTTCTTCAAGCAGTCCGACCGCGCGCTTACACCTTGCACGG 1008
DB 962 GTTGTGCTCAAGTGTCTTACTGCCAGGCGCAAGGATCGTGGGCTCTACGTCTGCGCG 1021
QY 1009 GTTCAACCTCGTGGCGAGACCTACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1068
DB 1022 CGCGCAACTCG 1081

RESULT 6

AAS28907

ID AAS28907 standard; DNA; 12415 BP.

XX AAS28907;

XX 07-NOV-2001 (first entry)

DE Human immunoglobulin encoding genomic DNA SEQ ID No 269.

KW Immunoglobulin; signal transduction pathway protein; cancer; ds;

KW Antisense therapy; gene therapy; neurological disorder; renal disorder;

KW Cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;

KW Reproductive disorder; immune system disorder; proliferative disorder;

KW Muscular disorder.

XX Homo sapiens.

XX WO200155315-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001326.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216860P.

XX 11-JUL-2000; 2000US-0217487P.

XX 14-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225288P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226581P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0234999P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236328P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244647P.
PR 08-NOV-2000; 2000US-0244647P.
PR 08-NOV-2000; 2000US-0244647P.

PR 08-NOV-2000; 2000US-0245476P.
PR 08-NOV-2000; 2000US-0245477P.
PR 08-NOV-2000; 2000US-0245478P.
PR 08-NOV-2000; 2000US-0245523P.
PR 08-NOV-2000; 2000US-0245524P.
PR 08-NOV-2000; 2000US-0245525P.
PR 08-NOV-2000; 2000US-0245526P.
PR 08-NOV-2000; 2000US-0245527P.
PR 08-NOV-2000; 2000US-0245528P.
PR 08-NOV-2000; 2000US-0245532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249298P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 03-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WFI; 2001-457725/49.

Isolated novel immunoglobulin polypeptide for monitoring the presence and progression of diseases and for diagnosis.
Claim 1; SEQ ID NO 269; 551bp; English.
Sequences AAS28878-AAS28926 represent genomic DNA molecules which encode the immunoglobulin polypeptides of the invention. The polynucleotides and polypeptides can be used to diagnose a pathological condition or a susceptibility to a pathological condition in a subject by determining the presence or absence of a mutation in a DNA sequence or determining the presence or amount of expression of the protein. Alternatively the identification of a binding partner to a sequence allows determination of changes in protein activity. The sequences can be used as research tools for receptors or other signal transduction pathway proteins that interact with the polypeptides of the invention and can be used to treat, prevent or diagnose various types of disorders such as neurological disorders, cardiovascular disorders, gastrointestinal disorders, reproductive disorders, immune system disorders, renal disorders, muscular disorders, pulmonary disorders, proliferative disorders and cancer. Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 12415 BP; 2667 A; 3521 C; 3458 G; 2768 T; 0 U; 1 Other;

Query Match 18.5%; Score 197.6; DB 4; Length 12415;
Best Local Similarity 51.7%; Pred. No. 4.7e-23;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

```
QY 7 CCAGCAGCCACACTCCGCGCGCTGCCAGAGCCGCCATAGAGAGAGTCCCGCCGCCA 66
Db 2 CCGTCCGACCCCGCTGCCCGCCACCGCTGCGCGCTGCGAGCGTCCGCGCGCGA 61
QY 67 CCGTATGATCAGCAGTCAAGCGGGGCGCGCTTCTCAGCGGGCCCAAGGCT 126
Db 62 TGAAGCGAGCTCGGGGATCAGGGAGCGCCCGCTTCTCGCTTCCCGCGGCTG 121
QY 127 TCGTGTGTGCGTGGGCAAGCAGCCCTCAGTGTCCAGATCGTGGTAATCCACGC 186
Db 122 TCGGGTGTAAAGTGGCGCCGAGCGAGCTCAAGTGGTCTCTGGGGAGCGCGCG 181
QY 187 CACAGTGTGAGTGGAGAGAACAGCAGCGGTGAAGCGCGCGCGCTTCCGTCTGG 246
Db 182 CTGTAGTGTGTGGAGAGAGCGGCGAGCAGCTGGCGCTCGGAACGCTGAGCTTC 241
QY 247 CCCAGGAGCGGACCTCTACCGCTCACTATCTCTGGACCTGGCGCTGGGACAGTGGC 306
Db 242 CGCGGAGCGCGGAGCAGCGCTGTGTGACCGCGGCACTGCCACGACGCGGGG 301
QY 307 AATAGTGTGCGCGCGCGCAATGCGATAGCGAGGCTTGTGCTGCGTGGGCTGCGAG 366
Db 302 TCTAGTGTGCGCGCGCGCAACCGCGCGGAGGCTACCGCGCGCGCGCTCACCG 361
QY 367 TGGAGCGGAGCGCGCTGCGCGAGAGCGCGCTTCTCTGCTGCGCGCGCGCTCCA 426
Db 362 TGCTGAGCGCGCGCTTCCAGCCGAGCTGACGCGCGGAGCGCGCTGCTCATCGC 421
QY 427 TCCGCGTGGCGAGGCTCAGAGCCACCTTCCGCTGCGG-----CGTGGGTGGCTCC 480
Db 422 CGGGTCCGCGAGGCGCGCGCTTCTCTCAGCGGCGCTCGATCCAGTGGGTGCTG 481
QY 481 CGAGCGCGAGTGTGCTGCTCAAGAGCGCGCGCGCTGCTGAGTCCGAGCGCGCGC 540
Db 482 GGGGCGGAGGTGTGCTGCTGCTGCTGCGCGCGCGCGCTTCCCGAGCGCGCGCTTACT 541
QY 541 CGTGTGCGGTGGAGAGTCTCGCGGAGGCAAGTGGCTGCGCATTCGGCGCGCGCGCG 600
Db 542 GGGAGAGGAGGATGCGCTTGACGAGAGTGTGGACAGCAGCACTTCGCGCTCCAGC 601
QY 601 CGAGCGCGGCACTTACAGTGTGCGCGCGAGAACCGCTGGGG--GCTGCCAGCGCGC 658
Db 602 CGGGCGCGCGAGGAGCGCGCGCGCGCGCTTGGCACTTGGCACTTGGCGGCTCGGC 661
QY 659 CGCGCGCTAGTG--GTGAGTCTGAGACCGCGGAGACCGGCGCGCGCGCGCGCTCCA 717
Db 662 TGCCGGATTCCGGGTCTAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 721
QY 718 CGCGCGCGCTTCTGCGGCACTTGTGAGCGCGCGCGAGGCTATGCGCGCGCGCGCGCG 777
Db 722 GGGCGCTGCTCCAGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
QY 778 CCGCTC-----ACCGCGCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828
Db 782 CCGCGCGGTGGTGGAGCGCTCAAGTGGCGCGCTAAGACCTTCTGGGTGAACGAGGCA 841
QY 829 AGCAGCGCGCTCAGTGTGCTGAGTACCGCGGAGCGCGCGCGCGCGCGCGCGCGCG 888
Db 842 AGCAGCGCAAGTTCCGCTGTGCTGATGGGAGCGCGCGCGCGCGCGCGCGCGCGCG 901
QY 889 AGGAGCGCGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948
Db 902 GGGAGGCGCGCGCTGCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 961
```

```
QY 949 ACTTCGTGTCAAGATCTCTTCTGCAAGCAGTCCGACCGCGCTTACACCTGCACGG 1008
Db 962 GTTTCGTGTCAAGGTCTTTACTGCCAGGCCAAGATCGTGGCTTACGTCTGCGCG 1021
QY 1009 CPTCCAACTCTGCGGCGAGCCTACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1068
Db 1022 CGGCAACTCGCGGCGCGAGCCTCAGTGGCGCTGACGTGCTGACGTGAAAGTACGCGG 1081
```

RESULT 7

```
AAI03356
ID AAI03356 standard; DNA; 12415 BP.
XX
AC AAI03356;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 6044.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 28-JUN-2000; 2000US-0209467P.
PR 30-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 26-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225575P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226886P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0227009P.
PR 01-SEP-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
```


PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231143P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0231141P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236328P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246603P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249218P.

PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-465570/50.
DR
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.
PT
XX
XX Disclosure; SEQ ID NO 6044; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention
XX
XX
XX Sequence 12415 BP; 2667 A; 3521 C; 3458 G; 2768 T; 0 U; 1 Other;
SQ

Query Match 18.5%; Score 197.6; DB 4; Length 12415;
Best Local Similarity 51.7%; Pred. No. 4,7e-23;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCAGAGCCACACTCCGCGCTGCCAGAGCCGCCATAGAGAGAGGTCGCCGCGCCA 66
Db 2 CCTCCGACCCCTCTGCCGCCACCGCTTCGCGCTGCGAGGGTCCGCGCGCGCA 61
QY 67 CCGTCATGGATCAGCCACAGTTTCAGCGGGGCGCGCGCTTCTACCCGCGCAAGGCT 126
Db 62 TGAAGCGAGCTCGGGGGATCAGGGAGCCCGCTTCCTTCGCGCTTCGCGCGCGCTG 121
QY 127 TCGTGTGTGGTGGGCAAGGAGCCACCTCAGCTGCCAGATCGTGGTAAATCCACGC 186
Db 122 TCGGGGTGTAAGTGGCGCGGAGGCGAGCTCAAGTGCCTGCTTGGGGAGCGCGCCG 181
QY 187 CACAGGTGAGCTGGGAGAAAGGACACAGCAGCCGCTGACGGCGCGCGCTTCCTCGT 246
Db 182 CTGTAGTGTGGTGGGAGAGGGCGGCGAGCTGGCGGCTCGGAACGCTGAGCTTC 241
QY 247 CCGAGAGCGGAGCTCTTACCGGCTCACTATCTGTGACCTGGGCTGGGCGAGTGGGC 306
Db 242 CGCGGAGCGCGGAGCAGCGGCTCTGCTGCTACCGCGCACTGCGCCACACGCGGGGG 301
QY 307 AATACGTGTCCGCGCGCAATGCCATAGGAGGAGGCTTCGCTGCGGTGGGCTGCAGG 366
Db 302 TCTACGTGTCCGCGCGCGGAGGAGGCTTACGGGGGGCGCGCTGAGCTTCC 361
QY 367 TGGACCGGAGGCGCGTTCGCGGAGCAGGCGCGCACTTCTGTGTGGGCGCCACGTCCA 426
Db 362 TGCTGAGCGCGCGGCTCTCGACCCGAGCTGAGCCCGCGAGCGCGCGCTGCCATCGC 421
QY 427 TCCGCGTGGCGAGGCGCTCAGAGGCCACCTTCGCTGCGG-----CGTGGGTGGTCCC 480

Db 422 CGGGTCCGGAGGGCCCCCGGCTTCTCCACAGGGGCTTCGATCCAGTGGGTGCTGC 481
QY 481 CGAGCCCGGAGTGAAGTGGTCCAGGACGGGGCGGCGCTGGGTGAGCCGACGGCCCCC 540
Db 482 GGGGGGCGAGTGGTGTGCTGACGTCCGGGGCGGGGGGCTCCCGAGGCCACACTGACT 541
QY 541 GGTTCGGCTGAGAGAGTCCGGCAGGCAAGTGGCTGCGCATTCGGGGCGGGCGGCCG 600
Db 542 GGGGAAGGACGGGATGCGCCCTGACGAGAGTGTGGGACAGCAGCCACTTCGCGCTCCAGC 601
QY 601 GCGACGGCGGCACTTACAGGTCCGCGCGGAGAACCCGCTGGGC - GCTGCGAGCGCGC 658
Db 602 CGGGCGGCGGAGGACGGCCCCGGCGGAGCTGGCACTGGCATCTCTGGCGCTCGGC 661
QY 659 CGGGCGGCTAGTG-GTGGACTCGGACGCGCGGACACGGCGAGCCCGCGGACTTCCA 717
Db 662 TCGCGGATTCGGGGCTTACGTGTGCCACGCCCGCCGCAACGGCGACGGCGCGG 721
QY 718 CGGGCGGCTCTGCGGACCTGACGGCGGCGGCGGAGGTATGCGCGCGAGGGCGCC 777
Db 722 GGGCGTCTCCAGTGCACGAGCCCCCGGAGAGCCCGCGGAGCCCGGAGGAGGCC 781
QY 778 CCGGCTC-----ACCGCCAGCACCGGCGGAGGCTGACCGGTGACTGAAGGCA 828
Db 782 CCGCGCGGTGGAGCGGCTCAAGTGGCGGCTAAGACCTTCTGGGTGACGAGGCA 841
QY 829 AGCAGCGGCGCTCAGCTGCTAGCTGACGGCGGAGCCAGCCGAGACGGTGTGAAGA 888
Db 842 AGCAGCGCAAGTTCCGCTGCTAGCTGATGGGCAAGCCGAGCCGAGATCGAATGGCACT 901
QY 889 AGGACGGCGGCTGCTGACGAGGCGCGGCGGCGGCTGACGAGGCGGCGGAGGCA 948
Db 902 GGGAGGCGGCGGCTGCTCCGAGCGCGGCGGCTTATGACCGGCGGCGGAGCGG 961
QY 949 ACTTGTGCTCAAGATCTCTTTCGAGGAGTGGCGGCGGCTTACCTGACGCG 1008
Db 962 GCTTCGTGCTCAAGTGTCTTACTGCCAGGCGGAGGATCGTGGGCTTACTCTGGCGG 1021
QY 1009 GGTCAACCTCTGCGGCGGAGGCTTACAGCTGTGTGGTGGTGTGCGGAGCGCGG 1068
Db 1022 CGCGCAACTCGGCGGCGGAGGCTGAGTGGCGGCTGACGTGACGTGAAGGTACGCGG 1081

RESULT 8

ADB31748

ID ADB31748 standard; DNA; 12415 BP.

AC ADB31748;

XX DT 04-DEC-2003 (first entry)

XX DE Human novel protein DNA SEQ ID NO 269.

XX KW Gene therapy; human; immunoglobulin; cancer; lung cancer; leukaemia;

XX KW ovarian cancer; epithelial cancer; hyperplasia; Gaucher's disease; AIDS;

XX KW arrhythmia; cardiac oedema; ischaemia; pneumonia; cystic fibrosis;

XX KW asthma; sarcoidosis; rhinitis; anaemia; inflammation; sinusitis;

XX KW chronic obstructive pulmonary disease; infectious disease; gene; ds.

XX OS Homo sapiens.

XX PN US2003077606-A1.

XX PD 24-APR-2003.

XX PF 07-MAR-2002; 2002US-00091438.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0225759P.
PR 22-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232403P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0234233P.
PR 21-SEP-2000; 2000US-0234234P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.

PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0241637P.
PR 01-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249224P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249266P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250300P.
PR 01-DEC-2000; 2000US-0250312P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251865P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764879.
(HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-615993/58.
XX
XX New human immunoglobulin superfamily of polypeptide and genes, useful for
PT treating, preventing or diagnosing e.g. cancers (lung cancer, leukemia,
PT ovarian epithelial cancer, etc.), hyperplasia, ischemia, pneumonia or
PT AIDS.
XX
XX
XX Disclosure; SEQ ID NO 269; 213pp; English.

CC The invention relates to an isolated polypeptide, which comprises the
CC human immunoglobulin superfamily of proteins. The polypeptide or
CC polynucleotide is useful for treating, preventing or ameliorating a
CC medical condition e.g. cancers (lung cancer, leukaemia, ovarian
CC epithelial cancer, etc.), hyperplasia, Gaucher's disease, AIDS,
CC arrhythmia, cardiac oedema, ischaemia, pneumonia, cystic fibrosis,
CC asthma, sarcoidosis, rhinitis, anaemia, leukaemia, inflammations,
CC sinusitis, chronic obstructive pulmonary disease, infectious diseases
CC etc. The polypeptide or polynucleotide is also useful for diagnosing any
CC of these diseases or a susceptibility to the disease. The present data for
CC sequence represent a novel human protein DNA. Note: The sequence data for
CC this patent did not form part of the printed specification but was
CC obtained in electronic format direct from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030077606
XX
SQ Sequence 12415 BP; 2667 A; 3521 C; 3458 G; 2768 T; 0 U; 1 Other;

Query Match 18.5%; Score 197.6; DB 9; Length 12415;
Best Local Similarity 51.7%; Pred. No. 4.7e-23;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCAGCAGCCACACTCCGGCCGCTGCCAGAGCCCCCATAGAGAGAGTCCCGCGCCCA 66
DB 2 CCTCGCACCCCTCCCTGCCACCGTTGCGGCTGCGAGGGTCCGGCGCGCGCA 61
QY 67 CCGTCATGGATCAGCCACAGTTTCAGCGGGGCGCCCGCTTCTACCCGCGCCAAAGGCT 126
DB 62 TGAAGGCGAGCTCGGGGATCAGGGAGCCCCCGTGTCTTCTCGCTTCCCGCGCGCTG 121
QY 127 TCGTGTGTGCTGGGCAAGGAGCCCTCAGCTGCCAGATCGTGGTAATCCACGC 186
DB 122 TCGGGTGTAAAGTGGCGCCGAGCGGAGCTCAAGTGTCTGTCTCTGGGGAGCGCCGC 181
QY 187 CACAGGTGAGCTGGGAGAGGACACAGCAGCCGTTGACGCCCGCGCGCTTCCGCTCG 246
DB 182 CTGTAGTGTGTGGAGAGGGCGGCGACAGCTGGCGGCTCGGAACCTCGAGTTCC 241
QY 247 CCCAGACCGGACCTCTACCGCTCACTATCTTGACCTGGGCTGGGCGAGTGGSC 306
DB 242 CGCGGAGCGCGGAGCAGCGCTGTCTGCTACCGCGCACTGCCACCGAGCGGGG 301
QY 307 AATACGTGTCCGCGCGCAATGCGCATAGGCGAGGCTTCCGCTGGCGCTGCAAG 366
DB 302 TCTAGTGTGCGCGCCGCAACGCGGCGCGGAGGCTTACCGGGCGCGCTCACCG 361
QY 367 TGAACCGGAGCGCGTGGCGGAGGAGCGCGGACCTTCTCTGTGGGCGCACGTCGA 426
DB 362 TGTGAGCGCGCGCTCCGACCGCGAGCTGACGCCCGCGAGCGCGCTGCCATCGC 421
QY 427 TCCGCGTGCAGAGGCTCAGAGGACCTTCCGCTGCGG-----CGTGGTGGCTCC 480
DB 422 CGGGTCCGGAGGCGCGCGCTTCTCTCACGGGCTCGATCCCACTGGTGTCTGC 481
QY 481 CGAGCGCGCAGTGAAGTGGTCCAAAGGACGGCGGCGCTGGGTGAGCGCGACGGCCCC 540
DB 482 GGGGGCGAGGTGGTGTGAGTGTGCGCGGGCGGGGCTCCCGAGCCACACTGTACT 541
QY 541 GCGTGGCGTGGAGGAGCTCGGCGAGGCAAGTGGCTGGCGATTGGGGCGCGCGCGC 600
DB 542 GGGAGAGGACGGGATGGCCCTGGGACAGTGGGACAGAGCCACTTCGCGCTCCAGC 601
QY 601 GCGACCGCGGCACTTACGAGTCCGCGCGAGAACCCGCTGGGC--GCTGCCAGCGCGC 658
DB 602 CGGGCGCGCGGAGGAGCGCGCGCGAGCTGGCACTGGCATCTCTGGCGGTGCGC 661
QY 659 CGCGCGCTAGTG- GTGAGTCCGAGCGCGGACAGCGCGAGCGCGCGCGCGGACCTCA 717
DB 662 TGGCGGATTCGCGGCTTACGTGTGCACGCCCGCAACGCGCACCGCGCGCGAGCGG 721
QY 718 GCGCGCGCTCTTGGCGCACCTTGCAGCGCGCGGAGGCTATGCGCGCGCGGCGCGCC 777
DB 722 GGGCGCTGTCTCAGGTGACACCGCGCGAGAGCGCGCGCGCGGAGCGCGCGCGCGC 781

```
QY 778 CCGCTC-----ACCGCCAGCAGCCGCGCAGCTGACCGTGAAGGCA 828
Db 782 CCGCGCCGGTGGAGCCGCTCAAGTGCAGCCCTAAGACCTTCTGGGTGAACGAGGCA 841
QY 829 AGCAGCGCGCTCAGTGTACGTGACCGCGAGCCCAAGCCGAGACGGTGTGGAAGA 888
Db 842 AGCAGCCAAAGTTCGGTCTACGTGATGGCAAGCCCGAGCCGAGATGCATGSCACT 901
QY 889 AGGACGGCCACTGGTGACCGAGCGCGGCCACGCTGGTGTACGAGGACCGCAGAGA 948
Db 902 GGGAGGGCGCCCGCTGCTCCCGAGCCCGCGCCCTCATGTACCGGACCGGACGGCG 961
QY 949 ACTTCGTGCTCAAGATCCTCTCTGCAAGCAGTGGAGCCGCGCTCTACACCTGACGCG 1008
Db 962 GCTTCGTGCTCAAGTGTCTTACTGCGAGGCAAGGATCGTGGCTTACTGTGCGCG 1021
QY 1009 CGTCAACCTCGTGGGCGACACTTACAGCTGTGTGCTGGTGTGAGTGGCGAGCCCGCG 1068
Db 1022 CGCGCAACTCGCGGCGCAGACGCTCAGTGCCTGCGAGCTGCACGTGAAAGGTACGCGCG 1081

RESULT 9
ID AAK89083
XX AAK89083 standard; DNA; 20565 BP.
AC AAK89083;
DT 05-NOV-2001 (first entry)
XX Human digestive system antigen genomic sequence SEQ ID NO: 2659.
DE Human; digestive system antigen; gene therapy; cancer; appendicitis;
XX ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX Homo sapiens.
OS
XX WO200155314-A2.
PN
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US001324.
PF
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216890P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225256P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0227009P.
PR 01-SEP-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
```


KW reproductive disorder; immune system disorder; proliferative disorder;
KW muscular disorder.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200155315-A2.
XX
XX
PD 02-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US001326.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236272P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240860P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241825P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251866P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.

FR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-457725/49.
XX
XX Isolated novel immunoglobulin polypeptide for monitoring the presence and
XX progression of diseases and for diagnosis.
XX
XX Claim 1; SEQ ID NO 270; 551pp; English.
XX
XX Sequences AAS28878-AAS28926 represent genomic DNA molecules which encode
XX the immunoglobulin polypeptides of the invention. The polynucleotides and
XX polypeptides can be used to diagnose a pathological condition or a
XX susceptibility to a pathological condition in a subject by determining
XX the presence or absence of a mutation in a DNA sequence or determining
XX the presence or amount of expression of the protein. Alternatively the
XX identification of a binding partner to a sequence allows determination of
XX changes in protein activity. The sequences can be used as research tools
XX for receptors or other signal transduction pathway proteins that interact
XX with the polypeptides of the invention and can be used to treat, prevent
XX or diagnose various types of disorders such as neurological disorders,
XX cardiovascular disorders, gastrointestinal disorders, reproductive
XX disorders, immune system disorders, renal disorders, muscular disorders,
XX pulmonary disorders, proliferative disorders and cancer. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 20565 BP; 4414 A; 5760 C; 6155 G; 4236 T; 0 U; 0 Other;
XX
XX Query Match 18.5%; Score 197.6; DB 4; Length 20565;
XX Best Local Similarity 51.7%; Pred. No. 4.4e-23;
XX Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;
XX
QY 7 CCAGCAGCCACACTCGCGCGCTGCCAGAGCCGCCATAGAGAGGTCCCGCGCCCA 66
Db 20564 CCTTCGCGACCCCGCTGCGCCCGCCACCGTTCGCGCTGCGAGCGGTGCGCGCGA 20505
QY 67 CCGTCATGATPACGCCAGTTCAGCGGGGCGCCCGCTTCTCACCGCGCCCAAGCCCT 126
Db 20504 TGAAGGCGAGCTCGGGGATCAGGGGAGCCCGCGTCTTCTCGCTTCCCGCGGCTG 20445
QY 127 TCGTGGTGTGCGTGGGCGAGCCAGCCACCTCAGCTGCCAGATCGTGGGTAACTCCACGC 186
Db 20444 TCGCGGTGGTAAAGTGGCGCGCGAGCGCGAGCTCAAGTGGTGGTCTTGGGGAGCGCGCG 20385
QY 187 CACAGGTGAGCTGGGAGAGGACACAGCGCGGTGACCGCGCGCGCGCTTCCGTCTGG 246
Db 20384 CTGTAGTGGTGGGAGAGCGCGGCGAGCAGCTGCGCGCTCGAAGCGCTGAGCTTCC 20325
QY 247 CCGAGCGCGAGCTTCTACCGCTTCTATCTCTGAGCTGCGCGCTGGCGGCGAGTGGGC 306
Db 20324 CCGCGGAGCGCGGAGCAGCGCTTGTGTGACCGCGCGACTGCCACCGAGCGCGGG 20265
QY 307 AATAGCTGCGCGCGCGCAATGCCATAGCGAGGCTTCTGCTGCGTGGCGCTGCAGG 366
Db 20264 TCTAGCTGTGCGCGCGCGCAACCGCGCGCGAGGCTTACGCGCGCGCGCTGACCG 20205
QY 367 TGGACGCGGAGCGCGCTGCGCGCGAGCGCGCGCAATCTCTCTGCGCGCGCGCTTCCA 426
Db 20204 TGCTGAGCGCGCGCTTCCAGCCCGAGCTGCGAGCGCGCGCGCGCTGCGCATCGC 20145
QY 427 TCCCGCTGCGGAGGCTCAGAGCGCCACCTTCCGCTCGCG-----CGTGGGTGCTTCC 480
Db 20144 CGGGTCCGGGAGGCGCGCGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 20085
QY 481 CGAGCGCGGAGTGAAGTGGTGGTCAAGAGCGCGCGCGCGCTGGGTGAGCGCGCGCGCGCG 540
Db 20084 GGGGGCGGAGTGGTGTGCTGACGTGCCGGGCGGGGGGCTTCCCGGAGCGCGCGCGCTACT 20025

QY 541 GCGTGGCGGTGAGAGCTCGCGAGGCAAGTGCCTCGGATTCGGCGCGCGCGCGCGC 600
Db 20024 GCGAGAGGACGCGGATGCGCTTGGACGAAGTGTGGGACAGCAGCGCACTTTCGCTTCCAGC 19965
QY 601 GCGAGCGCGGCTTACGAGGTCCCGCGCGGAGAACCGCTTGGC--GCTGCCAGCGCGC 658
Db 19964 CCGGCGCGCGCGAGGAGCGCGCGCGCGCGAGCTGCGCATCTTGGCGGTTCGCG 19905
QY 659 CCGCGCGCTAGTG-GTGACTCGGACGCGCGGAGCACGCGCGCGCGCGCGCGCGCTTCCA 717
Db 19904 TCGCGGATTCGCGCGTCTACGTGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19845
QY 718 CCGCGCGCGCTTCTGCGCGACCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777
Db 19844 GCGCGCTGCTCCAGGTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19785
QY 778 CCGCGCTC-----ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828
Db 19784 CCGCGCGCGGTGCGCGCGCGCTCAAGTGGCGCGCTTCTTGGGTGACAGCGCGCA 19725
QY 829 AGCAGCGCGCGCTCAGCTGCTGCTAGTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 888
Db 19724 AGCAGCG 19665
QY 889 AGGAGCGCGCGCTGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948
Db 19664 GCGAGCG 19605
QY 949 ACTTGTGTCTCAAGATCTCTTCTTCTCAAGAGTTCGCGCGCGCGCGCGCGCGCGCTTAC 1008
Db 19604 GCTTGTGTCTCAAGGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 19545
QY 1009 GCTTCAACCTGCG 1068
Db 19544 CCGCGCAACTCG 19485

RESULT 11
AAL03357
ID AAL03357 standard; DNA; 20565 BP.
XX
XX AAL03357;
XX
XX 21-NOV-2001 (first entry)
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 6045.
XX
XX Human reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US0011339.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184564P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189974P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX
XX 18-APR-2000; 2000US-0198123P.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX 28-JUN-2000; 2000US-0214886P.
XX
XX 30-JUN-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216880P.
XX
XX 11-JUL-2000; 2000US-0217487P.
XX
XX 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234233P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241755P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.

PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254037P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.

Disclosure; SEQ ID NO 6045; 1297pp + Sequence Listing; English.

CC The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention

XX Sequence 20565 BP; 4236 A; 6155 C; 5760 G; 4414 T; 0 U; 0 Other;

Query Match

Best Local Similarity 18.5%; Score 197.6; DB 4; Length 20565;

Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY

7 CCAGCAGCCCACTCCGGCGGTGCCAGAGCCCATAGAGAGGTCCCGCCGCCA 66

| | | | | | | |
|------|--------------|---------------------|--------------------------|----------------------|-------------|------|
| 2 | CCCTCGCACCC | CCCTCGCGCTGCAGGCGGT | CGCGCGCGGA | 61 | | |
| 67 | CGGTCAATG | CAGCACAGTTT | CAGCGGCGCCCGCTTTCT | CACCGCGCCCAAGGCT | 126 | |
| 62 | TGAAGCGAG | CTCGGGGAT | CAGGGAGCCCGCTGCTT | CTCGGCTTCCCGCGCCTG | 121 | |
| 127 | TCGTGTGTC | GGGCAAGAG | CCACCTCAGCTGCCAGAT | TCGTGGTATCCACGC | 186 | |
| 122 | TGCGGTGT | TAATGTCGCG | CAGGCGAGCTCAAGTGC | TGTTCCTGGGGAGCCCGCGC | 181 | |
| 187 | CACAGTGA | CTGGGAGAAG | CACAGCGGTGACGCGCGCGCGCTT | CCGTCTCTGG | 246 | |
| 182 | CTGTAGTGT | GGGAGAAGG | CGGCGAGAGCTGCGGCTTCG | NAACGCTGAGCTTCC | 241 | |
| 247 | CCAGAGCG | CGACCTTACCGCTCACTAT | CTGTGACCTGTGGCTGGGCGAC | AGTGGGC | 306 | |
| 242 | CGGCGAC | CGCGGAGCACGCGCTCTG | CTGACCGCGCACTGCCAC | CGACGCGGGG | 301 | |
| 307 | AATACCTG | TGCCCGCGCAATGCCAT | AGGAGAGCTTCGTGCGTGGGCT | TCGACG | 366 | |
| 302 | TCACGTGT | TGCCCGCGCAACGCGGCG | CGGAGGCTTACGCGGCGCGCT | CACCG | 361 | |
| 367 | TGACCGG | AGGCGCGGTGCGCGAG | CGCGCACTTCTGTGCGCGCCAC | GTCCA | 426 | |
| 362 | TGCTGAG | CGCGGCTCCGACCCGAG | CTGCAGCCCGCGCGCTGCCAT | CGC | 421 | |
| 427 | TCCCGTGC | GCAGGGCTCAGAGCCACTT | CCGCTGCCG----- | CGTGGTGGCTCC | 480 | |
| 422 | CGAGGT | CCGAGGCGCGCGCTT | CTCACAAGGCGCTCGAT | CCAGTGGTCTGC | 481 | |
| 481 | CGAGGCGG | CAGTAGCTGTGTCAAG | AGCGGCGCGCTTGGT | TGAGCCCGCGGCC | 540 | |
| 482 | GGGGGCG | AGGTGTGTGACGTG | CGGGGGGGGCTTCCCGAG | CCCACTGTACT | 541 | |
| 541 | GCGTGCG | CTGGAGAGCTCGGAGGCA | AGTGCCTGGGCA | TTCGGGCGCGCGGCC | 600 | |
| 542 | GGGAGAAG | ACGGATGGCCCTGGA | CGAAGTGGGACAGCAGCA | CTTCGCGCTCCAGC | 601 | |
| 601 | GCACGCG | CGCATTACGAGGTCCGCG | CGAGAACCGCTGGG-- | GCCTCAGCGCGCG | 658 | |
| 602 | CGGCGCG | CGCGAGGACGCGCGCGCG | AGCTTGGC | ACTTCTGGCGGCTCGGC | 661 | |
| 659 | CGCGGCG | CTAGTG-GTGGACT | TCGACCGCGGACAGCG | CACCGCGCGCGGACCTCCA | 717 | |
| 662 | TGCGGAT | TCGCGCTTACGTGTG | CACGCCCGCACCGCGCG | CGACGCGCGCG | 721 | |
| 718 | CGCGCGCG | CTCTGGCGCACTGTAG | CGGCGCGCGAGGCTAT | TGGGCGCGGCGCGCC | 777 | |
| 722 | GGGCGCT | TGCTCCAGGTTCACAG | CCCCCGCGCGGAC | CCCCCGCGGACCGCGCGC | 781 | |
| 778 | CGCGCTC----- | ACCGCCAGCACCGGAC | CGCACCTG | CACGCTGACTGAAGGCA | 828 | |
| 782 | CCGCGCG | GTGTGGAGCCGCTC | AGTGC | CGGCTTAAGACTTCTGGGT | GAACGAGGGCA | 841 |
| 829 | AGCACGCG | CGCCTCAGCTGTACT | CGTACCGCGGAGCC | CAAGCCCGAGACGGT | GTGGAAGA | 888 |
| 842 | AGCACGCA | AGTTCCGCTGTACGT | GATGGCAAGCCCGAC | CGCGAGATCGAAT | TGGCACT | 901 |
| 889 | AGGACGCG | CAGCTGGTACCGAGG | CGCGCGCACCTGTGT | TACGAGGACCGCGCAG | GAG | 948 |
| 902 | GGGAGG | CGCGCGCTCTCCGAC | CGCGCGCGCTCA | TGTACCGCGAC | CGCGAGCGCG | 961 |
| 949 | ACTTCTGT | CTCAAGATCTTCTG | CAGAGTCCGACCGCGGCT | CTACCTGTCA | CGG | 1008 |
| 962 | GCTTCTGT | CTCAAGTCTTACT | TGCCAGGCCAAGGAT | CTGTGGGCTCTA | CGTCTGCGCG | 1021 |
| 1009 | CGTCCAA | CTCGTGGGCGACCTT | CAGCTCTGTGTGTGT | GTGTGCGGAG | CGCGCGCG | 1068 |
| 1022 | CGCGCAA | CTCGCGGCGCAG | CGCTCAGTGCCTG | CTGACGCTG | CAAGGTA | 1081 |

| | |
|------------|--|
| ADB31749/C | ADB31749 standard; DNA; 20565 BP. |
| XX | |
| XX | ADB31749; |
| XX | |
| DT | 04-DEC-2003 (first entry) |
| XX | |
| DE | Human novel protein DNA SEQ ID NO 270. |
| XX | |
| KW | gene therapy; human; immunoglobulin; cancer; lung cancer; leukaemia; |
| KW | ovarian cancer; epithelial cancer; hyperplasia; Gaucher's disease; AIDS; |
| KW | arthritis; cardiac edema; ischaemia; pneumonia; cystic fibrosis; |
| KW | asthma; sarcoidosis; rhinitis; anaemia; inflammation; sinusitis; |
| KW | chronic obstructive pulmonary disease; infectious disease; gene; ds. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | US2003077606-A1. |
| XX | |
| PD | 24-APR-2003. |
| XX | |
| PF | 07-MAR-2002; 2002US-00091438. |
| XX | |
| PR | 31-JAN-2000; 2000US-0179065P. |
| PR | 04-FEB-2000; 2000US-0180628P. |
| PR | 24-FEB-2000; 2000US-0184664P. |
| PR | 02-MAR-2000; 2000US-0186350P. |
| PR | 16-MAR-2000; 2000US-0189874P. |
| PR | 17-MAR-2000; 2000US-0190076P. |

| | | |
|------------|----|---|
| ADB31749/C | ID | ADB31749 standard; DNA; 20565 BP. |
| XX | AC | ADB31749; |
| XX | AC | |
| XX | DT | 04-DEC-2003 (first entry) |
| XX | DE | Human novel protein DNA SEQ ID NO 270. |
| XX | XX | gene therapy; human; immunoglobulin; c |
| KW | XX | ovarian cancer; epithelial cancer; hyp |
| KW | XX | arthritis; cardiac oedema; ischaemia; |
| KW | XX | asthma; sarcoidosis; rhinitis; anaemia; |
| KW | XX | chronic obstructive pulmonary disease; |
| XX | XX | |
| OS | XX | Homo sapiens. |
| XX | XX | |
| XX | PN | US2003077606-A1. |
| XX | XX | |
| XX | PD | 24-APR-2003. |
| XX | XX | |
| XX | PF | 07-MAR-2002; 2002US-00091438. |
| XX | XX | |
| XX | XX | 31-JAN-2000; 2000US-0179065P. |
| XX | XX | 04-FEB-2000; 2000US-0180628P. |
| XX | XX | 24-FEB-2000; 2000US-0184664P. |
| XX | XX | 02-MAR-2000; 2000US-0186350P. |
| XX | XX | 16-MAR-2000; 2000US-0198874P. |
| XX | XX | 11-MAR-2000; 2000US-0190076P. |
| XX | XX | 18-APR-2000; 2000US-0198123P. |
| XX | XX | 19-MAY-2000; 2000US-0209467P. |
| XX | XX | 28-JUN-2000; 2000US-0214886P. |
| XX | XX | 30-JUL-2000; 2000US-0215135P. |
| XX | XX | 07-JUL-2000; 2000US-0216802P. |
| XX | XX | 11-JUL-2000; 2000US-0217487P. |
| XX | XX | 11-JUL-2000; 2000US-0217496P. |
| XX | XX | 26-JUL-2000; 2000US-0218293P. |
| XX | XX | 26-JUL-2000; 2000US-0220963P. |
| XX | XX | 14-AUG-2000; 2000US-0224518P. |
| XX | XX | 14-AUG-2000; 2000US-0224519P. |
| XX | XX | 14-AUG-2000; 2000US-0225213P. |
| XX | XX | 14-AUG-2000; 2000US-0225214P. |
| XX | XX | 14-AUG-2000; 2000US-0225266P. |
| XX | XX | 14-AUG-2000; 2000US-0225267P. |
| XX | XX | 14-AUG-2000; 2000US-0225268P. |
| XX | XX | 14-AUG-2000; 2000US-0225270P. |
| XX | XX | 14-AUG-2000; 2000US-0225447P. |
| XX | XX | 14-AUG-2000; 2000US-0225757P. |
| XX | XX | 14-AUG-2000; 2000US-0225758P. |
| XX | XX | 14-AUG-2000; 2000US-0225759P. |
| XX | XX | 18-AUG-2000; 2000US-0226279P. |
| XX | XX | 22-AUG-2000; 2000US-0226681P. |
| XX | XX | 22-AUG-2000; 2000US-0226868P. |
| XX | XX | 22-AUG-2000; 2000US-0227182P. |
| XX | XX | 23-AUG-2000; 2000US-0227009P. |
| XX | XX | 30-AUG-2000; 2000US-0228924P. |
| XX | XX | 01-SEP-2000; 2000US-0229287P. |
| XX | XX | 01-SEP-2000; 2000US-0229343P. |
| XX | XX | 01-SEP-2000; 2000US-0229344P. |
| XX | XX | 01-SEP-2000; 2000US-0229345P. |
| XX | XX | 05-SEP-2000; 2000US-0229509P. |
| XX | XX | 05-SEP-2000; 2000US-0230153P. |
| XX | XX | 06-SEP-2000; 2000US-0230437P. |
| XX | XX | 06-SEP-2000; 2000US-0230438P. |
| XX | XX | 08-SEP-2000; 2000US-0231242P. |
| XX | XX | 08-SEP-2000; 2000US-0231243P. |
| XX | XX | 08-SEP-2000; 2000US-0231244P. |
| XX | XX | 08-SEP-2000; 2000US-0231413P. |
| XX | XX | 08-SEP-2000; 2000US-0231414P. |
| XX | XX | 08-SEP-2000; 2000US-0232080P. |

```
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232406P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239933P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241212P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246612P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 08-SEP-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764879.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM, Barash SC;
XX WPI; 2003-615993/58.
DR
XX
XX New human immunoglobulin superfamily of polypeptide and genes, useful for
PT treating, preventing or diagnosing e.g. cancers (lung cancer, leukemia,
PT ovarian epithelial cancer, etc.), hyperplasia, ischemia, pneumonia or
PT AIDS.
XX
XX Disclosure; SEQ ID NO 270; 213pp; English.
PS
XX
XX The invention relates to an isolated polypeptide, which comprises the
CC human immunoglobulin superfamily of proteins. The polypeptide or
CC polynucleotide is useful for treating, preventing or ameliorating a
CC medical condition e.g. cancers (lung cancer, leukemia, ovarian
CC epithelial cancer, etc.), hyperplasia, Gaucher's disease, AIDS,
CC arrhythmia, cardiac oedema, ischaemia, pneumonia, cystic fibrosis,
CC asthma, sarcoidosis, rhinitis, anaemia, leukaemia, inflammations,
CC sinusitis, chronic obstructive pulmonary disease, infectious diseases
CC etc. The polypeptide or polynucleotide is also useful for diagnosing any
CC of these diseases or a susceptibility to the disease. The present
CC sequence represent a novel human protein DNA. Note: The sequence data for
CC this patent did not form part of the printed specification but was
CC obtained in electronic format direct from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030077606
XX
XX Sequence 20565 BP; 4414 A; 5760 C; 6155 G; 4236 T; 0 U; 0 Other;
SQ
Query Match 18.5%; Score 197.6; DB 9; Length 20565;
Best Local Similarity 51.7%; Pred. No. 4.4e-23;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;
QY 7 CCAGAGCCACATCCGGCCGCTGCCAGAGCCGCCATAGAGAGGTCCCGCCGCCA 66
Db 20564 CCCTCCGACACCCCTGCCCCCCTGCGGCTGCGGCTGCTCCGCGGCTG 20505
QY 67 CCGTCATGATCAGCCACAGTTACGCGGGCGCCCGCTTTCTCACCGGCGCCAAAGGCGCT 126
Db 20504 TGAAGGCGAGCTCGGGGATCAGGGGAGCCCGCGTCTTCTCGCTTCCCGGCGCTG 20445
QY 127 TCGTGGTGTGCGTGGGCAAGGACCCACCTCAGCTGCGAGATGTGGGTAATCCACGC 186
Db 20444 TCGGGTGTAAAGTGGCGCGGAGCCGAGCTCAAGTGGTGGTCTGCGGGAGCGCGCG 20385
QY 187 CACAGGTGAGCTGGGAGAAGGACAGCAGCGCGGTGAGCGCGGCGCGCTTCGCTCTGG 246
Db 20384 CTGTAGTGTGTGGGAGAAAGGGCGGCGAGAGCTGGGCGGCTCGGAACGCTGAGCTTCC 20325
QY 247 CCCAGGACGGGCACTCTACCGCTCTACTATCTCTGGAACCTGGGCGCTGGGCGAGTGGGC 306
Db 20324 CGGCGGACGGCGGAGCAGCGGCTGCTGTGTACCGCGGCACTGCCACCGCGCGGGG 20265
QY 307 AATACGTGTGCGCGCGCGCAATGCCATAGCGAGGCTTTCGCTGCGCTGGGCTGCAGG 366
Db 20264 TCTACGTGTGCGCGCGCGCGCAACCGCGCGCGGAGGCTACGCGGCGCGCGCGCGG 20205
```

QY 367 TGGACGGAGCGCGCTGCGCGAGCGCGCGCACTTCTGCTCGCGCCACGCTCCA 426
Db 20204 TGTGGAGCGCGCGCTCGACCGCGCGCTGCGAGCGCGCGCGCTGCGCATCGC 20145
QY 427 TCCGCGTGGCGAGCGCTCAGAGCGCGCGCTTCCGCTCGCG-----CGTGGGTGGCTCC 480
Db 20144 CGGGTCCGGGAGGGCGCGCGCTTCTTCTTACGGGGCTTGAATCCAGTGGTGTGCTG 20085
QY 481 CGAGCGCGGAGTGAAGTGTGCTTCAAGGACGGCGCGCGCTTGGGTGAGCCCGAGCGCGCC 540
Db 20084 GGGGGCGGAGGTGGTGTGCTGACGTGCGCGCGCGGGCGCGCTTCCCGAGCGCCACACTGTACT 20025
QY 541 CGTGGCGGTGAGAGCTGGCGAGGAGGAGTGGCTGCGCATTCGGGCGCGCGCGCGCG 600
Db 20024 GGGAGAGGACGGATGGCTGACCGGAGGAGTGTGGGACAGCGCACTTGGCGCTCCAGC 19965
QY 601 CGACGGCGGCACTTACGAGTCCGCGCGCGAGAACCGCGCTGGGG--GCTGCCAGCGCGCG 658
Db 19964 CGGGCGCGCGAGGAGCGCGCGCGCGCGCGCTTGGCATCTTGGCGGTGGC 19905
QY 659 CGCGCGGTAGTG--GTGAGTTCGACGCGCGCGGAGCACGGCGCGCGCGCGCGCGCTTCCA 717
Db 19904 TCGCGGATTCGGGGGTCTACGTGTGCCACCGCGCGCGCGCGCGCGCGCGCGCG 19845
QY 718 CGGCGCGCTCTTGGCGCACTGCGAGCGCGCGCGCGCGCGCTTATCGCGCGCGAGGCGCG 777
Db 19844 GGGCGCTCTCAGTGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19785
QY 778 CGCGCTC-----ACG 828
Db 19784 CGCGCGGTGTGGAGCGCGCTCAAGTGGCGCGCTTAAGCTTCTGGGTGAACGAGGCGCA 19725
QY 829 AGCAGCGCGCTCAGTGTGCTACGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 888
Db 19724 AGCAGCGCAAGTTCGCTGCTACGTGATGGGCAAGCGCGCGCGCGCGCGCGCGCG 19665
QY 889 AGGAGCGCGCGTGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948
Db 19664 GGGAGCGCGCGCGCTGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19605
QY 949 ACTTCGTGCTCAAGATCTTCTTCAAGCAGTGGGACCGCGCGCGCGCGCGCGCGCG 1008
Db 19604 GTTCGTGCTCAAGGTGCTTACTGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 19545
QY 1009 GTTCAACCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1068
Db 19544 CCGCAACTCG 19485

RESULT 13
ID ADD14722 standard; cDNA; 5382 BP.
XX
AC ADD14722;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human src biomarker polynucleotide SEQ ID NO:116.
XX
KW predictor set; protein tyrosine kinase activity modulator;
XX protein tyrosine kinase pathway; protein tyrosine kinase; cytoskeletal;
XX gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
OS
XX Homo sapiens.
XX
XX WO2003062395-A2.
XX
XX 31-JUL-2003.
XX
XX 17-JAN-2003; 2003WO-US001981.
XX
XX 18-JAN-2002; 2002US-0350061P.
XX

(BRIM) BRISTOL-MYERS SQUIBB CO.
Huang F, Fairchild CR, Lee FY, Shaw P;
WPI: 2003-636735/60.
P-PSDB; ADD14123.
New polynucleotides and polypeptides for predicting the activity of
compounds that interact with protein tyrosine kinases and/or protein
tyrosine kinase pathways.
Claim 2; SEQ ID NO 116; 139pp; English.
The present invention describes a predictor set comprising a plurality of
polynucleotides or polypeptides whose expression pattern is predictive of
the response of cells to treatment with a compound that modulates protein
tyrosine kinase activity or members of the protein tyrosine kinase
pathway. Also described: (1) predicting whether a compound is capable of
modulating the activity of cells, comprising obtaining a sample of cells,
determining whether the cells express a plurality of markers, and
correlating the expression of the markers to the compound's ability to
modulate the activity of the cells; (2) a plurality of cell lines for
identifying polynucleotides and polypeptides whose expression levels
correlate with compound sensitivity or resistance of cells associated
with a disease state; and (3) identifying polynucleotides and
polypeptides that predict compound sensitivity or resistance of cells
associated with a disease state, comprising subjecting the plurality of
cell lines to one or more compounds, analysing the expression pattern of
a microarray of polynucleotides or polypeptides, and selecting
polynucleotides or polypeptides that predict the sensitivity or
resistance of cells associated with a disease state by using the
expression pattern of the microarray. The polynucleotides and
polypeptides have cytostatic activities, and can be used in gene therapy.
The polynucleotides and polypeptides are useful in predicting the
activity of compounds that interact with protein tyrosine kinases and/or
protein tyrosine kinase pathways. These may be used in determining drug
sensitivity in patients to allow the development of individualized
genetic profiles which aid in treating diseases and disorders (e.g.
cancer) based on patient response at a molecular level. The present
sequence is used in the exemplification of the present invention.
Sequence 5382 BP; 943 A; 1699 C; 1798 G; 942 T; 0 U; 0 Other;
Query Match 11.8%; Score 126; DB 9; Length 5382;
Best Local Similarity 59.2%; Pred. No. 1.2e-11;
Matches 238; Conservative 0; Mismatches 155; Indels 9; Gaps 1;
QY 672 GTGACTCGAGCG 731
Db 125 GTCTACTGTGTGCG 184
QY 732 GCGCACCTGCG 784
Db 185 GTGACCGAGCG 244
QY 785 --ACCG 842
Db 245 GAGCGCGCTCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 304
QY 843 AGTGTCTACGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 902
Db 305 CGTGTCTACGTGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 364
QY 903 GTGACCGAGCG 962
Db 365 CTGCTCCG 424
QY 963 ATCTCTCTTGTCAAGCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1022
Db 425 GTGCTCTACTGCG 484
QY 1023 GCGCAGACCTTACAGCTCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1064

Db 6915 CCCCAGCTGGACGGACACCTGCACGCGACAGACCCCGCTTCGCCCTGCAGACCGCGCT 6974
QY 404 CTTCTCTGTGGCGGCCACAGCTCCATCCGGGTGGCGAGGGCTCAGAGGCCACTTTCGGCTG 463
Db 6975 GSCCGCCCAAGCTGGGGCGTTTCGGCTCTGGGCGGACCGGGTGGCGGCGCACTCGGTGG 7034
QY 464 CCGGTGGGTGCTCCCGAGCCCGGACGTAGCTGGTCCAGGACGGGCGGCGCTGGG 523
Db 7035 CGAGATCGCGCGCGGCGACGTCCGGGCGCGCTTCGCCCTCGACGCGGCGTGGCGCTGGT 7094
QY 524 TGAGCCGACGACCCCGGCTGGGTGGAGAGCTCGGCGAGGCAAGTGGCGCTGGCAT 583
Db 7095 CCGCAGCGCGGTGCTCATCGCGACCTGCGACCGGGCGGATGGTCCGCCGTGAGGT 7154
QY 584 TCGGGCGCGCGCGCGCGACCGCGCACTTA-----CGAGTCCGCGCGAGAA 634
Db 7155 CACCGAGGAGGCGCTGGACGCGATCGGCGACCTGCCGCTCGCGTTCGCCGTCTCAA 7214
QY 635 CCGGTGGCGGTGCCAGGCGCGCGCGCTAGTGTGACTCGGACGCGCGCGGACAC 694
Db 7215 CCGCGCGAGGCGGTGCTCTCCGCGACCGCGAGGCGGTCTGGCGCGTTCGGGCGCG 7274
QY 695 GGCACGCGCGCGCGGACCTCCACGCGCGCGCTCCTGGCGCACCTGCAGCGCGCGCGCGA 754
Db 7275 CCTCGCGCGCGCGCAAGCGGCTCCCGGTGGACCGGGGTTCCACTCGCGCGACGTGA 7334
QY 755 GGCTATGC---GCGCGAGGGCGCCCGCTTCACCGGCCAGCACCGGCACTG 811
Db 7335 CCGGTGCTCGACGCGCTGGCGGCTGGCCACCGGACCCCGACACACCCCTGATCTC 7394
QY 812 CACGTGTACTGAAGCAAGCACGCGCTCAGCTGTACTGACCGCGCGAGCCCAAGCC 871
Db 7395 CACCAACCGCGCGACGCTGACCCACGTCGGCCACACCACTGGGCGCGACCGCGG 7454
QY 872 CGAGACGTGTGGAAAGGACGCGCCAGCTGGTGACCGAGGGCGCGGCCACGTTGGTGA 931
Db 7455 CGGCGGCTCCGCTTCGCGGACGCGCTGCGCGCCCTGGCGGCGACGCGCGTCAAGCGG 7514
QY 932 CGAGACGCGCAGGAGAACTTGTGTCAAGATCCTTCTGCAAGCAGTGGACCGCGG 991
Db 7515 CGTGAGCTGGGCGCGCGCGTGTCAACCGCGACCGCGCGAGTCTTGGCGCGCGC 7574
QY 992 CCTTACACCTGCACGCGGTCCAACTCGTGGGCGGACCTACAGCTCTGTGCTGGTGGT 1051
Db 7575 GATCCCGCGCTGCGCGCGGACCAAGAGGTGCTGCGCTGCTCACCGCGCTGGCGGT 7634
QY 1052 AGTGGCGGAGCCCG 1065
Db 7635 GGTGCACGCGCGG 7648

Search completed: September 19, 2004, 15:42:33
Job time : 448.18 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:09:24 ; Search time 96.0922 Seconds
(without alignments)
6179.453 Million cell updates/sec

Title: US-10-077-130-4_COPY_1_1070

Perfect score: 1070
Sequence: 1 tgcctaccagcagccacac.....tagtgccgagccgcgggtt 1070

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|--------------------|-------------------|
| C 1 | 68.6 | 6.4 | 1614 | US-09-616-289-45 | Sequence 45, Appl |
| C 2 | 68.6 | 6.4 | 12425 | US-08-616-289-50 | Sequence 50, Appl |
| 3 | 65 | 6.1 | 3575 | US-09-023-655-1395 | Sequence 1395, Ap |
| 4 | 64.8 | 6.1 | 44377 | US-08-804-227C-7 | Sequence 7, Appli |
| 5 | 64.8 | 6.1 | 44377 | US-08-804-198-1 | Sequence 1, Appli |
| C 6 | 62.6 | 5.9 | 8438 | US-07-945-283-1 | Sequence 1, Appli |
| C 7 | 62.6 | 5.9 | 4403765 | US-09-103-840A-2 | Sequence 2, Appli |
| 8 | 61.8 | 5.8 | 1191 | US-09-046-922-3 | Sequence 3, Appli |
| 9 | 61.8 | 5.8 | 1389 | US-08-484-438-41 | Sequence 41, Appl |
| 10 | 61.8 | 5.8 | 1797 | US-08-463-163-2 | Sequence 2, Appli |
| 11 | 61.8 | 5.8 | 1839 | US-09-479-479-1 | Sequence 1, Appli |
| 12 | 61.8 | 5.8 | 1839 | US-09-297-851-1 | Sequence 1, Appli |
| 13 | 61.8 | 5.8 | 1869 | US-08-356-786-15 | Sequence 15, Appl |
| 14 | 61.8 | 5.8 | 2760 | US-08-743-637B-20 | Sequence 20, Appl |
| 15 | 61.8 | 5.8 | 2760 | US-08-526-840B-20 | Sequence 20, Appl |
| 16 | 61.8 | 5.8 | 2760 | US-08-748-170A-3 | Sequence 3, Appli |
| 17 | 61.8 | 5.8 | 2760 | US-09-047-148-1 | Sequence 1, Appli |
| 18 | 61.8 | 5.8 | 5727 | PCT-US91-02954-13 | Sequence 13, Appl |
| 19 | 61.6 | 5.8 | 4403 | US-08-284-941-1 | Sequence 1, Appli |
| 20 | 61.6 | 5.8 | 4403 | US-08-447-642-1 | Sequence 1, Appli |
| 21 | 61.6 | 5.8 | 4403 | US-09-236-503-1 | Sequence 1, Appli |
| 22 | 61.6 | 5.8 | 4403 | PCT-US93-02147A-1 | Sequence 1, Appli |
| 23 | 61.2 | 5.7 | 4257 | US-08-690-473-1 | Sequence 1, Appli |
| 24 | 61.2 | 5.7 | 4257 | US-09-259-821A-1 | Sequence 1, Appli |
| 25 | 61.2 | 5.7 | 4257 | US-08-843-659-1 | Sequence 1, Appli |
| C 26 | 61 | 5.7 | 2561 | US-09-616-289-48 | Sequence 48, Appl |
| 27 | 60.8 | 5.7 | 2185 | US-08-173-508-3 | Sequence 3, Appli |

| | | | | | | |
|------|------|-----|---------|---|---------------------|-------------------|
| 28 | 60.8 | 5.7 | 2185 | 2 | US-08-265-310-3 | Sequence 3, Appli |
| 29 | 60.8 | 5.7 | 2185 | 3 | US-08-951-742-3 | Sequence 3, Appli |
| 30 | 60.4 | 5.6 | 1029 | 4 | US-09-252-991A-135 | Sequence 135, App |
| 31 | 60.4 | 5.6 | 2457 | 4 | US-09-252-991A-120 | Sequence 120, App |
| C 32 | 60.4 | 5.6 | 2707 | 2 | US-08-709-874A-9 | Sequence 9, Appli |
| C 33 | 60.4 | 5.6 | 2707 | 2 | US-09-104-382-9 | Sequence 9, Appli |
| C 34 | 60.4 | 5.6 | 2707 | 4 | US-09-833-553-9 | Sequence 9, Appli |
| C 35 | 60.2 | 5.6 | 2001 | 4 | US-09-252-991A-3177 | Sequence 3177, Ap |
| 36 | 60.2 | 5.6 | 2012 | 1 | US-08-235-838-13 | Sequence 13, Appl |
| 37 | 60.2 | 5.6 | 2012 | 1 | US-08-235-838-15 | Sequence 15, Appl |
| 38 | 60.2 | 5.6 | 2012 | 2 | US-08-465-473B-13 | Sequence 13, Appl |
| 39 | 60.2 | 5.6 | 2012 | 2 | US-08-465-473B-15 | Sequence 15, Appl |
| 40 | 60.2 | 5.6 | 2304 | 4 | US-09-252-991A-2790 | Sequence 2790, Ap |
| 41 | 60.2 | 5.6 | 2561 | 4 | US-09-616-289-48 | Sequence 48, Appl |
| 42 | 59.8 | 5.6 | 77536 | 4 | US-09-410-551B-1 | Sequence 1, Appli |
| 43 | 59 | 5.5 | 1187 | 1 | US-08-440-856A-2 | Sequence 2, Appli |
| 44 | 58.6 | 5.5 | 1908 | 3 | US-09-046-992-1 | Sequence 1, Appli |
| C 45 | 58.6 | 5.5 | 4411529 | 3 | US-09-103-840A-1 | Sequence 1, Appli |

ALIGNMENTS

RESULT 1
US-09-616-289-45/c
; Sequence 45, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1614)
US-09-616-289-45

| | | | | |
|-----------------------|--------|--|-------|-----------------|
| Query Match | 6.4% | Score 68.6; | DB 4; | Length 1614; |
| Best Local Similarity | 45.7%; | Pred. No. 0.00029; | | |
| Matches | 274; | Conservative | 0; | Mismatches 324; |
| | | | | Indels 1; |
| | | | | Gaps 1; |
| Qy | 238 | TCGCTGTGGCCAGGACGGGACCTTACCGGCTCACTATCTTGACCTGGCTGGCGG 297 | | |
| | | | | |
| Db | 863 | TCCAGACGGCCTCGAGCCGCCCTCTCTCTCAGCAGCCCTGACCGCGCGGGGTT 804 | | |
| | | | | |
| Qy | 298 | ACAGTGGCAATACGTGTGCCGCGCGCAATGCCATAGCGGAGGCTTCGCTGCCGTGG 357 | | |
| | | | | |
| Db | 803 | AGCGACCGCGCGCGCGCGGTGCTCCCGAGGTAGCGACGACTTCCCGCAGGCTCAG 744 | | |
| | | | | |
| Qy | 358 | GCTGTGAGTGTGACGCGGAGCGCGGTGCGCGAGCAGCGCGCGGCACTTCCTGTGCGGC 417 | | |
| | | | | |
| Db | 743 | GGCG 684 | | |
| | | | | |

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: 9402646
US-09-023-655-1395

Query Match 6.1%; Score 65; DB 4; Length 3575;
Best Local Similarity 51.6%; Pred. No. 0.0012;
Matches 149; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 79 AGCCACAGTTCAGCGGGCGCCCGCTTCTCACCGGCCCAAGCCCTTGGTGTGCGG 138
Db |||||

Qy 139 TGGGCAAGGAGCCACCTCAGTCCGACATGCTGGGTAAATCCACGCCACAGTGAAGT 198
Db |||||

Qy 3204 CTGGGTACTCGGCGCCCTCACTGTGTGTGACAGAGGCCACCCGAAAGCGGTGTCT 3263
Db |||||

Qy 199 GGGAGAGGACACAGCCGCTGAGCGCGCGCGCTTCCGTCCTGGCCACGAGCGCG 258
Db |||||

Qy 3264 GGATGAAGAACAGATGGAATCCGTGAAGATCCCAAGTTCTGATAACCAATTACCAAG 3323
Db |||||

Qy 259 ACCTCTACCGCTCACTATCTCGACCTTGGGCGGCGGCGGCGGCGGCGGCGGCGG 318
Db |||||

Qy 3324 GAGTCTGAGCGCTGAACATCGTCCGCCCTTCGAGGCTGGGACTTACACCTGCC 3383
Db |||||

Qy 319 GCGGCGGCAATGCCATAGGCGAGGCTTCGCTGCGGTGGGCGCTGCAGGT 367
Db |||||

Qy 3384 GGGCGGTCAACGAGCTGGGCGAGGCGCTGGGTGAGTGCAAGCTGGAGGT 3432
Db |||||

RESULT 4

US-08-804-227C-7

Sequence 7, Application US/08804227C

Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: DeHoff, Bradley S.

APPLICANT: Kuhstoss, Stuart A.

APPLICANT: Rostock, Paul R., Jr.

APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

STREET: THOMAS G. PLANT 1501

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: USA

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII(DOS) text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804,227C

FILING DATE: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 4377 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

Query Match 6.1%; Score 64.8; DB 2; Length 44377;
Best Local Similarity 46.7%; Pred. No. 0.0013;
Matches 382; Conservative 0; Mismatches 422; Indels 14; Gaps 5;

Qy 222 AGGCGCGCGCGCGCTTCGGTCTGGCCACGAGCGGAGCTCTACCGCTCACTATCTCTG 281
Db |||||

Qy 23568 ACGGATGGCGGCGCTTTCGGCGCGCGGCTCTGCGGAAACCGCGGAGTCCCGCGCC 23627
Db |||||

Qy 282 GACCTGGCGCTGGGCGACAGTGGGCAATACGTGTGCGCGCGCGCAATGCCATAGCGGAG 341
Db |||||

Qy 23628 GGGCTGGCGCGCGCGGCGACCGTCTCTATCGCGCGGCGACCTCACAGC----GTGCC 23683
Db |||||

Qy 342 GCGTTCGCTGCGCGTGGCGCTGACGTGGACGCGGAGGCGCGTGGCGCGGAGCGGCGCG 401
Db |||||

Qy 23684 GCGGACTGGTCCGCTCCCTCTCGAGGACGCGCGCGGACCGCGTGGTGTCTGGCGGAC 23743
Db |||||

Qy 402 CACTTCTCTGCTGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 461
Db |||||

Qy 23744 AGGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 23802
Db |||||

Qy 462 TGCGCGGTGGTGGTTCCTCCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 521
Db |||||

Qy 23803 TGCGAGGTCAACG 23862
Db |||||

Qy 522 GGTGAGCGCGAGCG 581
Db |||||

Qy 23863 GCGGTGACG 23922
Db |||||

Qy 582 ATTGCGGCG 641
Db |||||

Qy 23923 ATCG 23979
Db |||||

Qy 642 GCGCTGCG 701
Db |||||

Qy 23980 CCG 24039
Db |||||

Qy 702 CGGCG 761
Db |||||

Qy 24040 GCGGCG 24099
Db |||||

Qy 762 CG 818
Db |||||

Qy 24100 CG 24159
Db |||||

Qy 819 ACTGAAGG---CAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 875
Db |||||

Qy 24160 GCGGCG 24219
Db |||||

Qy 876 ACGGTGTGAAGAAGGACG 935
Db |||||

Qy 24220 CAGGCGGTGCGGCG 24279
Db |||||

Qy 936 GAGCGCGCAGGAGAACTTCGTGTCTCAAGATCTCTTCTGCAAGCAGTTCGAGCGCG 995
Db |||||

Qy 24280 GACGTGAGTGGAGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24339
Db |||||

QY 996 TACACCTGCACGGCTCCAACTCTGTGGGCGACACCTA 1033
DB 24340 GACGACATCCCGACGCTGACGGCTGCGCGCGCCGA 24377

RESULT 5
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rosteck, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
US-08-804-198-1

Query Match 6.1%; Score 64.8; DB 2; Length 44377;
Best Local Similarity 46.7%; Pred. No. 0.0013;
Matches 382; Conservative 0; Mismatches 422; Indels 14; Gaps 5;

QY 222 ACGCGCGCGCGCGCTCCGCTGTGGCCAGACGGCGACTCTACCGCTCTCATATCTGT 281
DB 23568 ACGGATGGCGCGCTCTCGCGCGCGCGCTCTCGGAAACCGCGCGACTCTCCCGCGCCC 23627

QY 282 GACCTGGCGTGGCGACAGTGGGCAATACGTGTGCGCGCGCGCAATGCCATAGGCGAG 341

DB 23628 GGCCTGGCGCGCGCGGCGACCGTCTCTCATCGCGCGGACCTCAGACG-----GTGCGCG 23683
QY 342 GCCTTCGCTGCGGTGGGCTGCAGGTGGACGGGAGSCCGGTGGCGGAGCAGGCGCG 401
DB 23684 GCCGACTGGTCCGCTCCTCTCAGGACGGCGGACCGGCTGGTGTGCGCGGACCG 23743
QY 402 CACTTCTCTGCTGCGCGCCACCGTCCATCCGCTGGCGGAGGCTCAGAGGCCACCTTCGCG 461
DB 23744 ACGCCCCCGCACAGACGCGCG-CCGCGGACTGACCGGCGTCTCCCTGCTCCCGTGGCG 23802
QY 462 TCCCGCTGGTGGTGGTCTCCCGAGGCGGCACTGAGTGTGTCAGAGGAGGGCGGCGCTG 521
DB 23803 TCGACGCTCAGCGACCGCGCGCACTGGCCCGCTGTCTGACGAGCAGCGGCCACCGCT 23862
QY 522 GGTGAGCCCGACGGCGCGCGCTGCGCTGGAGGAGCTCGCGGAGCAAGTGGCTGCGCG 581
DB 23863 GCCGTGACGCGCGCGCGCTGTGCTGCTGCGCGCTGGCGGAGCGGACCGCGCGAC 23922
QY 582 ATTGGGCGCGCGCGCGCGCGAGCGGCGCACTTACGAGGTCCGCGCGCGAGAACCGCTG 641
DB 23923 ATCGCGCGCGCGCTCGC---CGCCAAAGACCAAGCGCGCGCGCGCTGCTGCTGCG 23979
QY 642 GCGCTGCGACGCGCGCGCGCGCTAGTGTGAGCTCGAGCGCGCGGACACCGCGCGAC 701
DB 23980 CCGCGCGCGCGCTCGACGCGCTGTGCTGTCTCTCGCTCGAGTGTGGGGCGG 24039
QY 702 CGCGCGCGGACCTCCACGCGCGCGCTCTGTGCGCACTTGCAGCGCGCGCGCGAGGCTATG 761
DB 24040 GCGCGCGCGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCG 24099
QY 762 CGCGCGGAGGCG 818
DB 24100 CGCGCGCGGCGGTGCGCGCGCTTCTCGTGGCGCTGGAGCGCGCGCGCGCGCGCGCG 24159
QY 819 ACTGAAG---CAAGCAGCGCGCGCTCAGCTGTCTGTGAGCGCGCGCGCGCGCGCGCG 875
DB 24160 GCCGACGCTGCGGCGGAGTCTCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24219
QY 876 ACGGTGTGAGAGAGGAGCGCGCGCGCTGTGACCGAGGCGCGCGCGCGCGCGCGCGCG 935
DB 24220 CAGCGCGTGGCGACCGCTGCGCGCGCATGTGTGAGCGCGCGCGCGCGCGCGCGCG 24279
QY 936 GACGCGCAGGAGAACTTGTGCTCAAGATCTCTTCTGCAAGCAGTGGAGCGCGCGCGCT 995
DB 24280 GACGTGAGTGGAGCGGTTGCGCGCGCTCTACACCTGGGTGCGTCCCGCGTACTTCT 24339
QY 996 TACACCTGCACGGGTCCAACTCTGTGGGCGCAGACCTA 1033
DB 24340 GACGACATCCCGACGCTGACGGCTGCGCGCGCCGA 24377

RESULT 6
US-07-945-283-1/c
; Sequence 1, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EPO and LIT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL USA
; COUNTRY: IL USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/945,283
;; FILING DATE: 19920911
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Ribando, Curtis P
;; REGISTRATION NUMBER: 27976
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 309-685-4011 ext.513
;; TELEFAX: 309-685-4128
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 8438 base pairs
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Pseudorabies virus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 622..6495
;; FEATURE:
;; NAME/KEY: variation
;; LOCATION: replace(1099, "g")
;; FEATURE:
;; NAME/KEY: variation
;; LOCATION: replace(1267, "t")
;; FEATURE:
;; NAME/KEY: variation
;; LOCATION: replace(1381, "c")
;; FEATURE:
;; NAME/KEY: variation
;; LOCATION: replace(1566, "c")
;; FEATURE:
;; NAME/KEY: variation
;; LOCATION: replace(7010, "g")
US-07-945-283-1

Query Match 5.9%; Score 62.6; DB 1; Length 8438;
Best Local Similarity 42.8%; Pred. No. 0.0033;
Matches 384; Conservative 0; Mismatches 504; Indels 9; Gaps 1;

QY 182 CACGCCACAGGTGAGCTGGGAGAGGACACGACCGGTGACGCGCGCGCGCGCTTCG 241
DB 4834 CATGCCCTACCGGGCGCGCGCGGACCCCGCGCGGAGCGACGTCGAGGCCCT 4775

QY 242 TCTGGCCACGACCGCGACCTCTACCGCTCACTATCTGGACCTGGCGCTGGCGACG 301
DB 4774 CTGGCGCGCGCTCGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGAGCT 4715

QY 302 TGGGCAATACGTGCGCGCGCGCAATGCCATAGGCGAGCCTTCGCTGCGTGGCGCT 361
DB 4714 GCGGACCGCTGCTGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCT 4655

QY 362 GCAGTGGACCGGAGCGCGCTGCGCGAGCAGCGCGCGCACTTCCTGCTGGCGGCCAC 421
DB 4654 CCGGCTGCGAGGT 4595

QY 422 GTCCATCCCGTGGCGGAGGCTCAGAGCGCACCTTCGCTGCGCGCGCGCGCGCGCGCG 481
DB 4594 GACCGCGCGCTGCTGCGCGCTTCGCGAGCGGATCCCGGGGCGCGCGCGCGCGCGGCA 4535

QY 482 GAGGCGCGCAGTGAGCTGGTCCAGGACGCGCGCGCGCTGGGTGAGCCCGACGCGCGCG 541
DB 4534 GCAGGCGCGGACTCTGCTGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCTACAG 4475

QY 542 CGTGGCGTGGAGAGCTTCGCGGAGGCAAGTGCCTGCGCATTCGCGCGCGCGCGCGCG 601
DB 4474 CGTGACGGGCG 4415

QY 602 CGACGGCGGACATTACGAGGTCCGGCGCCGAGAACCCGCTGGGCGCTGCAGCGCGCGCG 661
DB 4414 GGCACACGTGGCGCGCGCGCGCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGAGCC 4355

QY 662 GCGCTAGTGGTGGACTCGACGCGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCG 721
DB 4354 GCGCGCGCGCTGCGCGCGCGCTGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCG 4295

QY 722 CCGCTCTTGGCGCACCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
DB 4294 GCG 4235

QY 782 CTCACCGCGCGACCG 841
DB 4234 CACCAAGTCCGGCTCCAGCACCAAGTCCAGCTCCGCGCGCGCGCGCGCGCGCT 4175

QY 842 CAGCTCTACGTGA-----CGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 892
DB 4174 CTCGCGCTACGCGAGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4115

QY 893 CGGCGAGCTGGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 952
DB 4114 GAAGCG 4055

QY 953 CGTGTCTCAAGATCCTCTTCTGCAAGCAGTCGCGCGCGCGCGCGCGCGCGCT 1012
DB 4054 CGGCTCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3995

QY 1013 CAACCTCTGGCG 1069
DB 3994 CCGGCGCAAGCGCGCGCTCCCTCGGACTCGGCGCGCGCGCGCGCGCGCT 3938

RESULT 7

US-09-103-840A-2/c

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 5.9%; Score 62.6; DB 3; Length 4403765;
Best Local Similarity 47.0%; Pred. No. 0.0032;
Matches 231; Conservative 0; Mismatches 254; Indels 6; Gaps 1;

QY 219 GTGACGGCGCGCGCGCGCTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 2059206
DB 2059265 GTACCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2059206

QY 279 CTGGACCTGGCGCTGGGCGACAGTGGGCAATACGTGTCGCGCGCGCGCAATCCATAGC 338
DB 2059205 GTGGCGCGGTGTTGGCGGTGACGCTGTGCGCATTCCTGGGACACCGCGCGCGCG 2059146

QY 339 GAGGCGCTTCGCTGCGCGCTGGGCGCTGAGGTGAGCGCGCGCGCGCGCGCGCG 398

| | | | | |
|----|--|---------|--|---------|
| Db | | 2059145 | GGTCCGGCGGGCCGGTGGCTGCTTCGGCGTCGGTGGGGCCGGCGCGCCGGCGGAA-- | 2059088 |
| QY | | 399 | CCGCACTTCGTGTCGGGCCACGTCCATCCGCGTGGGAGGCGCTAGAGGCCACTTC | 458 |
| Db | | 2059087 | ---TCGGA TTGGTTCGGAAACGGCGGTGTCGGGGGGTCCGGCGGTCCGCCCTGCTCTGG | 2059032 |
| QY | | 459 | CGTGCCGGTGGGTGGCTCCCGAGCGCGGCAGTGAGCTGGTCCAAGACAGGGCGGCGC | 518 |
| Db | | 2059031 | GGCGACGGCGGTGCCGGCGCGGGTGGGTCCGCTTCACTACCGCGGTGCCGGCGG | 2058972 |
| QY | | 519 | CTGGGTAGCCCGACCGCCCCCGCTGCGGTGGAGGAGCTCGGCGAGGCAATGTCGCTG | 578 |
| Db | | 2058971 | GCGGCGGCAACCCCGCCTTGCTGGTAGCGCGCGGGCGCGCGCGCGCGCACTC | 2058912 |
| QY | | 579 | CGCATTCGGCGCGCGCGCGCGCGACGGCGGCACCTTACGAGGTCCGCGCCGAGAACCGC | 638 |
| Db | | 2058911 | GGCGGTGGCGCTACCGGGGTGGCGCGCGCGCGCGGAAACGGCGGCATCGGGCGCTG | 2058852 |
| QY | | 639 | CTGGGCGCTCCAGCGCGCGCGCGCTAGTGGTGGACTCGAGCCCGCGACAGGCC | 698 |
| Db | | 2058851 | TTTGGTCCCGCGCGCGCGCGCGCGCGGCGGATTCGGCTTCGGCGGTCCCGGGGCGCC | 2058792 |
| QY | | 699 | AGCCGGCCCGG 709 | |
| Db | | 2058791 | GGTGGGCTCGG 2058781 | |

RESULT 8

US-09-046-992-3
; Sequence 3, Application US/09046992
; Patent No. 6140066

GENERAL INFORMATION:
APPLICANT: Lorberboum-Galski, Haya
APPLICANT: Yarkoni, Shai
APPLICANT: Ben-Yehudah, Ahni
TITLE OF INVENTION: METHODS OF CANCER DIAGNOSIS
TITLE OF INVENTION: USING A CHIMERIC TOXIN
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:

```

? ZIP: 10036-2911
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows
? SOFTWARE: FASTSEQ for Windows Version 2.0b
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/046.992
? FILING DATE: 24-MAR-1998
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
?
```

```

; LOCATION: 1...1198
; OTHER INFORMATION:
US-09-046-992-3

Query Match          5.8%; Score 61.8; DB 3; Length 1191;
Best Local Similarity 48.0%; Pred. No. 0.0046;
Matches 207; Conservative 0; Mismatches 222; Indels 2; Gaps 1

QY 434 GCGAGGGCTCAGAGCCACCTTCCTCCGCTCCGCTGGGTGGCTCCCGAGCGCCGGCAGT 493
DB 220 GTGCAGCGGCTGGTGCCTCTACCTGGCGCGCGCTGTCTGTGGAACCCAGGTCGACCAG 279

QY 494 GAGCTGTGTCAGAGACGGGCGGCGCTGGGTGAGCC--CGACGGCCCCCGCTGCGCGTG 551
DB 280 GTGATCCGACAGCCCTGGCGAGCCCGCGAGCGGCGGCGACCTGGGGAAGCATCGCG 339

QY 552 GAGGAGCTCGCGAGGCAAGTGCGCTCGCATTCCTGGGCGCGCGGCGCGAGCGCGCGC 611
DB 340 GAGCAGCGGAGCAGGCCCTCTGGCCCTGACCTGGCGCGCGCGAGCAGCGCGCTTC 399

QY 612 ACTTACGAGTTCGCGCGCGAGAACCGCTGGGCGCTCCAGCGCGCGCGCGCGTGTAGTG 671
DB 400 GTCCGCGAGGCAACCGGCAACACGAGGCGCGCGCGGCCAACCGCGAGTGTGTAGCCTG 459

QY 672 GTGGACTCGAGCGCGGGGACACGGCCAGCGCGCGCGGAGCTCCACGCGCCGCGTCTCG 731
DB 460 ACCTGCCCGTCCGCGCGGTTGAATGCGCGGCGCGCGGACAGCGGCGACGCCCTGTG 519

QY 732 GCGCACTGCAAGCGCGCGCGAGGCTATCGCGCGCGAGGGCGCCCGCGCTCACCGGCC 791
DB 520 GAGCGCAACTATCCCACTGGCGCGGAGTTCTCTGGGCAAGGGGGGACGTGAGCTTCAGC 579

QY 792 AGCACCGGCACGCGCACTGTCAGGTCGACTGAAGGCAAGCAGCGCGGCTTCAGCTGCTAC 851
DB 580 ACCCGGGCACGCAAGACTGGACGGTGGACCGGCTGCTCCAGCGCACCGCCAACTGGAG 639

QY 852 GTGACCGGCGA 862
DB 640 GAGCGCGGCTA 650

```

RESULT 9

US-08-484-438-41
; Sequence 41, Application US/08484438
; Patent No. 5811098
; Patent No. 5811098 5780031

? Patent No.: 3681028 3780031
 ? GENERAL INFORMATION:
 ?
 ? APPLICANT: Plowman, Gregory D.
 ? APPLICANT: Culouscou, Jean-Michel
 ? APPLICANT: Shoyab, Mohammed
 ? APPLICANT: Siegall, Clay B.
 ? APPLICANT: Hellstr m, Ingegerd
 ? APPLICANT: Hellstr m, Karl E.
 ? TITLE OF INVENTION: HB4 HUMAN RECEPTOR TYROSINE KINASE
 ? NUMBER OF SEQUENCES: 42
 ? CORRESPONDENCE ADDRESS:
 ?

```

? ZIP: 100362/1
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/484,438
? FILING DATE: 07-JUN-1995
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/323,442

```

```
; FILING DATE: 14-OCT-1994
; APPLICATION NUMBER: US 08/150,704
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/981,165
; FILING DATE: 24-NOV-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 5624-230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1389 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-484-438-41

Query Match          5.8%; Score 61.8; DB 1; Length 1389;
Best Local Similarity 48.0%; Pred. No. 0.0046;
Matches 207; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

Qy 434 GCGGAGGGCTCAGAGGCCACTTTCGCTGCGCGTGGTGGTTCCTCCCGAGCGCGCAGT 493
Db 418 GTGCAGCGGTGGTGGCTTACTTGGCGCGCGGTGCTGTTGAACACAGGTCGACCA 477
Qy 494 GAGCTGTGTCAGAGCAGCGCGCGCTGGGTGAGCC--CGACGCGCGCGCGTCCG 551
Db 478 GTGATCCGCAACGCCCTGGCGCAGCCCGCGGAGCGCGGCGGCGGAGCGATCCG 537
Qy 552 GAGGAGCTCGGAGCGGCAAGTGCCTGCGCATTCGGCGCGCGCGCGCGCGCGCGG 611
Db 538 GAGCAGCGGAGCAGGCGCGCTGTGCGCTGACCTTGGCGCGCGCGGAGCGCGTTC 597
Qy 612 ACTTACGAGTTCGCGCGCGAGAACCGGTGGGCGCTGCGAGCGCGCGCGCGCGCTAGT 671
Db 598 GTCCGCGCAGGCGACCGGCAACGAGCAGCGCGCGCGCGCGCGCGCGCGCGTGTGAGC 657
Qy 672 GTGAGCTCGGAGCGCGCGGACACGCGCAGCGCGCGCGCGCGCGCGCGCGCTCTG 731
Db 658 ACCTGCCGCTGCGCGCGGTGAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 717
Qy 732 GCGCACTTCAGCGCGCGCGCGAGGCTATGCGCGCGAGGCGCGCGCGCGCGCGCGCC 791
Db 718 GAGCGCAACTATCCCACTGGCGCGGAGTTCCTCGCGCAGCGCGCGCGCGCGCGT 777
Qy 792 AGCAGCGCAGCGCAGCTGACCGTGAAGGCGAGCAGCGCGCGCGCGCGCTAC 851
Db 778 ACCCGCGCAGCGCAGCACTGACCGTGAAGCGGTGAGCGCGTGTCTCCAGCGCGCG 837
Qy 852 GTGACCGCGA 862
Db 838 GAGCGCGGCTA 848

RESULT 10
US-08-463-163-2
; Sequence 2, Application US/08463163
; Patent No. 5696237
; GENERAL INFORMATION:
; APPLICANT: Fitzgerald, David J.
; APPLICANT: Chaudhary, Vijay K.
; APPLICANT: Pastan, Ira H.
```

```
; APPLICANT: Waldmann, Thomas A.
; APPLICANT: Queen, Cary L.
; TITLE OF INVENTION: Recombinant Antibody-Toxin Fusion Protein
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,163
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/227,227
; FILING DATE: 22-JAN-1981
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/911,227
; FILING DATE: 24-SEP-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,361
; FILING DATE: 21-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/865,722
; FILING DATE: 08-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 015280-12211
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1797
; OTHER INFORMATION: /product= "Anti-Tac(Fv)-PE40"
; US-08-463-163-2
```

```
Query Match          5.8%; Score 61.8; DB 1; Length 1797;
Best Local Similarity 48.0%; Pred. No. 0.0046;
Matches 207; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

Qy 434 GCGGAGGGCTCAGAGGCCACTTTCGCTGCGCGTGGTGGTTCCTCCCGAGCGCGCAGT 493
Db 829 GTGCAGCGGTGGTGGCTTACTTGGCGCGCGGTGCTGTTGAACACAGGTCGACCA 888
Qy 494 GAGCTGTGTCAGAGCAGCGCGCGCTGGGTGAGCC--CGACGCGCGCGCGTTCGCGTG 551
Db 889 GTGATCCGCAACGCCCTGGCGCAGCCCGCGGAGCGCGCGCGCGGAGCGATCCGC 948
Qy 552 GAGGAGCTCGGAGCGCAGTGGCTGCGCATTCGGCGCGCGCGCGCGCGCGCGCG 611
Db 949 GAGCAGCGGAGCAGCGCGCTGTGCGCTTACCTTGGCGCGCGCGCGCGCGCGCGTTC 1008
Qy 612 ACTTACGAGTTCGCGCGCGAGAACCGGTGGGCGGTGCGAGCGCGCGCGCGCGCTAGT 671
Db 1009 GTCCGCGCAGGCGACCGGCAACGAGCGCGCGCGCGCGCGCGCGCGCGCTGTGAGC 1068
Qy 672 GTGAGCTCGGAGCGCGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCTCTC 731
```


| | | | |
|----|------|--|------|
| Db | 1069 | ACCTGCCCGGTCCGCCCGGTGAATCGCGGGCCCGCGGACAGCGGCGCCCTGCTG | 1128 |
| QY | 732 | GCGCACCTGCAGCGGGCGCGAGGCTATGCGCGCGAGGGCGCCCGCCTCACGCGCC | 791 |
| Db | 1129 | GAGCGCAACTTATCCCACTGCGCGGAGTTCTCGGCGACGGCGGCGACGTCAAGCTTCAGC | 1188 |
| QY | 792 | AGCACCGGACGCGCACCTGCACGGTGACTGAGGCAAGCACGCGCGCTCAGCTGCTAC | 851 |
| Db | 1189 | ACCGCGGACGCGAAGTGAACGGTGGAGCGCGTCTCAGGCGCACGCCAACTGGAG | 1248 |
| QY | 852 | GTGACCGCGA | 862 |
| Db | 1249 | GAGCGCGCTA | 1259 |

```

RESULT 11
US-09-479-479-1
/ Sequence 1, Application US/09479479
/ Patent No. 6423513
/ GENERAL INFORMATION:
/ APPLICANT:
/ APPLICANT:
/ APPLICANT:
/ TITLE OF INVENTION: Protease-Activatable Pseudomonas
/ TITLE OF INVENTION: Exotoxin A-Like Proteins
/ NUMBER OF SEQUENCES: 36
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Townsend and Townsend and Crew LLP
/ STREET: Two Embarcadero Center, Eighth Floor
/ CITY: San Francisco
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94111-3834
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/479,479
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 09/297,851
/ FILING DATE: 30-JUL-1999
/ APPLICATION NUMBER: US 60/030,376
/ FILING DATE: 06-NOV-1996
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: WO PCT/US97/20207
/ FILING DATE: 05-NOV-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Storella, John R.
/ REGISTRATION NUMBER: 32,944
/ REFERENCE/DOCKET NUMBER: 015280-29810US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 576-0200
/ TELEFAX: (415) 576-0300
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1839 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..1839
/ OTHER INFORMATION: /product= "Pseudomonas exotoxin
/ OTHER INFORMATION: A"
US-09-479-479-1

```

| | | | | |
|-----------------------|--------------|-------------------|-----------------|--------------|
| Query Match | 5.8%; | Score 61.8; | DB 4; | Length 1839; |
| Best Local Similarity | 48.0%; | Pred. No. 0.0046; | | |
| Matches 207; | Conservative | 0; | Mismatches 222; | Indels 2; |
| Gaps | 1; | | | |

| | | | |
|----|------|--|------|
| QY | 434 | GGCGAGGGCTCAGAGGCGACCTTCGGCTGCGCGCTGGGTGCTCCCGAGGCGGCGACT | 493 |
| DB | 871 | GTGACGGCTGGTGGCTTACTTGGGGCGCGCTGTCGTGAAACGAGTCCAGCAG | 930 |
| QY | 494 | GAGCTGGTCAAGAGCAGGCGCGCGCTGGGTGAGCC--CGACGGCCCCCGCGTGC | 551 |
| DB | 931 | GTGATCGCAACGCCCTGSCCAGCCCCGCGCAGCGCGGCGACTGGGCGAAGCATCCGC | 990 |
| QY | 552 | GAGGAGCTCGGAGGCAAGTCCGTGCGCATTCGGGCGGCGCGCGCGCGACGCGCGC | 611 |
| DB | 991 | GAGCAGCGGAGCAGGCGCGCTGTGCGCTGACCCCTGGCGCGCGCGCGAGCGCGCTTC | 1050 |
| QY | 612 | ACTTACGAGGTCGGCGCGAGAACCCGCTGGGCGCTGCAGCGCGCGCGCGCGCTAGTG | 671 |
| DB | 1051 | GTCCGCGAGGCAACCGGCAACGAGAGCGCGCGCGGCGCAACGCGAGCGTGTGAGCCTG | 1111 |
| QY | 672 | GTGGAATGGAGCGCGGGAACAGGCGCAGCGCGCGCGCGGAACTTCACGGCGCGCTCTG | 731 |
| DB | 1111 | ACCTGCCCGTGGCGCGCGGTGAATGCGCGGCGCGCGCGGCGAGCGCGCGCTGCTG | 1170 |
| QY | 732 | GGCGACCTGCAGCGCGCGCGAGGCTATCGCGCGCGAGGGCGCCCCCGCTCACCGCC | 791 |
| DB | 1171 | GAGCGCACTATCCACTGGCGGAGTTCTCGCGAGCGCGCGCGCGAGCTCAGCTTCAGC | 1230 |
| QY | 792 | AGCACCGGACGCGCACTGTGACGCTGACTGAGGCAAGCAGCGCGCGCTCAGCTGCTAC | 851 |
| DB | 1231 | ACCGCGGCGCAGAACTGGACGCTGGAGCGGTCTCCAGGCGCACCGCGCACTGGAG | 1290 |
| QY | 852 | GTGACCGGCGA | 862 |
| DB | 1291 | GAGCGCGGCTA | 1301 |


```

RESULT 12
US-09-297-851-1
; Sequence 1, Application US/09297851
; Patent No. 6426075
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: Protease-Activatable Pseudomonas
; TITLE OF INVENTION: Exotoxin A-Like Proteoproteins
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/297,851
; FILING DATE: 30-JUL-1999
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,376
; FILING DATE: 06-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/20207
; FILING DATE: 05-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John R.

```

```

;
;   REGISTRATION NUMBER: 32,944
;   REFERENCE/DOCKET NUMBER: 015280-29810US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 576-0200
;   TELEFAX: (415) 576-0300
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1839 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..1839
;   OTHER INFORMATION: /product= "Pseudomonas exotoxin A"
;
; US-09-297-851-1
;
; Query Match          5.8%; Score 61.8; DB 4; Length 1839;
; Best Local Similarity 48.0%; Pred. No. 0.0046;
; Matches 207; Conservative 0; Mismatches 222; Indels 2; Gaps 1;
;
; QY 434 GCGGAGGGCTCAGAGCCACCTTCCGCTGCGGCTGGGTGGCTCCCGAGGCGGCAGT 493
; DB 871 GTGACGGGCTGGTGCCTTACTTGGCGGGCTGTCTGTGGAAACCAAGTGCACCG 930
; QY 494 GAGCTGTGTCGAAGACGGGCGGCGCTTGGGTGAGCC--CGACGCGCCCGCGTGGCGGTG 551
; DB 931 GTGATCCGCAACGCCCTGGCCAGCCCGGCGAGCGGCGGACCTGGGCGAAGCGATCCGC 990
; QY 552 GAGAGCTGCGGAGGCAAGTGCCTGGGCAATTGCGGCGGGCGGCGCGCGAGCGGCG 611
; DB 991 GAGCAGCGGAGAGCGGCGGCTGTGGCCCTGACCCCTGGCCCGCGCGGAGAGCGGCGTTC 1050
; QY 612 ACTTACAGGTCGCGCGCGAGAACCCGCTGGGCGCTGCGAGCGCGCGCGCGCTAGTG 671
; DB 1051 GTCCGGAGGCGACCGGCAAGCAGAGCGCGCGCGCGCGCGCGCGCGTGTGAGCTG 1110
; QY 672 GTGACTGCGAGCGCGGCGACACGCGCAGCGCGCGCGGCGCGCGCGCGCGCGCGCTCTG 731
; DB 1111 ACCTGCCCGTGTGCGCGCGCGTGAATGTGGCGGCGCGCGCGCGCGCGCGCGCGCTG 1170
; QY 732 GCGCACCTGACGCGCGCGGAGGCTATGCGCGCGCGGCGCGCGCGCGCGCGCTACCGCCC 791
; DB 1171 GAGCGCAACTATCCACTGCGCGCGGAGTTCCTCGCGACGCGCGCGCGCGCTCAGCTT 1230
; QY 792 AGCACCGGACGCGACCTGACCGGTGACTGAAGCGAAGCAGCGCGCGCGCTCAGCTGTAC 851
; DB 1231 ACCCGCGCAGCGAGAACTGGACGCTGGAGCGGCTGCTCCAGGCGCACCGCGCAACTGG 1290
; QY 852 GTGACCGCGGA 862
; DB 1291 GAGCGCGGCTA 1301
;
; RESULT 13
; US-08-356-786-15
; Sequence 15, Application US/08356786
; Patent No. 5877305
; GENERAL INFORMATION:
; APPLICANT: Huston, James S.
; APPLICANT: Oppermann, Hermann
; APPLICANT: Houston, L. L.
; APPLICANT: Ring, David B.
; TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA

```

```

;
;   ZIP: 02109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC Compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/356,786
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/831,967
;   FILING DATE: 06-FEB-1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Pitcher, Edmund R.
;   REGISTRATION NUMBER: 27,829
;   REFERENCE/DOCKET NUMBER: CRP-053
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 248-7000
;   TELEFAX: (617) 248-7100
;   INFORMATION FOR SEQ ID NO: 15:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1869 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..1869
;   OTHER INFORMATION: /note= "Product = "741sFv-PE40"
;
; US-08-356-786-15
;
; Query Match          5.8%; Score 61.8; DB 2; Length 1869;
; Best Local Similarity 48.0%; Pred. No. 0.0046;
; Matches 207; Conservative 0; Mismatches 222; Indels 2; Gaps 1;
;
; QY 434 GCGGAGGGCTCAGAGCCACCTTCCGCTGCGGCTGGGTGGCTCCCGAGGCGGCAGT 493
; DB 898 GTGACGGGCTGGTGCCTTACTTGGCGGGCGGCTGTCTGTGGAAACCAAGTGCACCG 957
; QY 494 GAGCTGTGTCGAAGACGGGCGGCGCTTGGGTGAGCC--CGACGCGCCCGCGTGGCGGTG 551
; DB 958 GTGATCCGCAACGCCCTGGCCAGCCCGGCGAGCGGCGGCGACTGGGCGAAGCGATCCGC 1017
; QY 552 GAGGAGCTGCGGAGGCAAGTGGCTGGCAATTCGGGCGGCGCGCGCGCGCGCGCGCG 611
; DB 1018 GAGCAGCGGAGCAGGCGCGCTGTGGCCCTGACCCCTGGCGCGCGCGCGCGCGCGCTT 1077
; QY 612 ACTTACAGGCTGCGCGCGCGAGAACCCGCTGGGCGCTGCGAGCGCGCGCGCGCGCTAGTG 671
; DB 1078 GTCCGCGAGGCGACCGGCAACGACGAGGCGCGCGCGCGCGCGCGCGCGCTGAGCCTG 1137
; QY 672 GTGGAATCGGACGCGCGGACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTG 731
; DB 1138 ACCTGCCCGTGTGCGCGCGTGAATGTGGCGGCGCGCGCGCGCGCGCGCGCGCTGCTG 1197
; QY 732 GCGCACCTGACGCGCGGCGCGAGGCTATGCGCGCGCGAGGCGCGCGCGCGCGCTCACCG 791
; DB 1198 GAGCGCAACTATCCACTGCGCGGAGTTCCTCGGCGACGCGCGCGCGCGCTCAGCTT 1257
; QY 792 AGCACCGGACGCGCACCTTGCACGCTGTAAGGCAAGCAGCGCGCGCGCTCAGCTGTAC 851
; DB 1258 AACCGCGGACGCGAGAACTGGACGCTGGAGCGGCTGCTCCAGGCGCGCACCGCACTGG 1317
; QY 852 GTGACCGCGGA 862
; DB 1318 GAGCGCGGCTA 1328
;
; RESULT 14
; US-08-743-637B-20
; Sequence 20, Application US/08743637B

```

Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743.637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2760 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
US-08-743-637B-20

Query Match 5.8%; Score 61.8; DB 2; Length 2760;
Best Local Similarity 48.0%; Pred. No. 0.0046;
Matches 207; Conservative 0; Mismatches 222; Indels 2; Gaps 1;
QY 434 GCGGAGGGCTCAGAGCCACCTTCGCTCCGCGTGGTGGCTCCCGAGGCGGCAGT 493
DB 1691 GTGAGCGGCTGTGCGCCCTTACCTGGCGCGGCTGTGGAACAGGTGACACG 1750
QY 494 GAGTGTGTCAAGAGACGGGCGGCGCTGGGTGAGCC--CGACGGCCCCCGCGTGGCGTG 551
DB 1751 GTGATCCGCAACGCCCTTGGCGAGCCCGCGGCGGCGACCTGGCGGAAGCATCCG 1810
QY 552 GAGAGCTCCGCGAGGCAAGTGGCTCGCTATTCGGCGCGCGCGCGCGACGCGCG 611
DB 1811 GAGAGCGGAGAGAGGCGGCTGTGGCCCTGACCTGGCGCGCGCGAGAGCGAGGCTTC 1870
QY 612 ACTTACAGGTCCGCGCGAGAACCCGCTGGGCGCTCCAGCGCCCGCGCGGCTAGTG 671
DB 1871 GTCCGGCAGGCGCACCGGCAACGACGAGCGCGCGGCGGCGCAAGCCGACGTGTGAGCTG 1930
QY 672 GTGAGTCCGAGCGCGGACAGCGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 731
DB 1931 ACCTGCCCGGTGCGCGCGGTGAATGCGCGGCGCGGCGGCGGCGGCGGCGGCGG 1990
QY 732 GCGCACTGCGAGCGGCGCGGAGGCTATGCGCGCGCGGCGGCGGCGGCGGCGGCGG 791

DB 1991 GAGCGCAACTATCCCACTGGCGGAGTTCTCGCGGAGCGGCGGCGGCGGCGG 2050
QY 792 AGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 851
DB 2051 ACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2110
QY 852 GTGACCGGCGG 862
DB 2111 GAGCGGCGGCTA 2121
RESULT 15
US-08-526-840B-20
Sequence 20, Application US/08526840B
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2760 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
US-08-526-840B-20

Query Match 5.8%; Score 61.8; DB 3; Length 2760;
Best Local Similarity 48.0%; Pred. No. 0.0046;
Matches 207; Conservative 0; Mismatches 222; Indels 2; Gaps 1;
QY 434 GCGGAGGGCTCAGAGCCACCTTCGCTCCGCGTGGTGGCTCCCGAGGCGGCAGT 493
DB 1691 GTGAGCGGCTGTGCGCCCTTACCTGGCGCGGCGGCTGTGGAACAGGTGACACG 1750
QY 494 GAGTGTGTCAAGAGACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 551
DB 1751 GTGATCCGCAACGCCCTTGGCGAGCCCGCGGCGGCGGCGGCGGCGGCGGCGG 1810

| | | | |
|----|------|--|------|
| Qy | 552 | GAGAGCTCGCGAGGCAAGTGCCTCGCTCGCATTTGGCGCGCGCGCGCGCGACGCGGC | 611 |
| Db | 1811 | GAGCAGCGGAGCAGGCCGCTCTGGCCCTGACCTTGGCCCGCGCGAGACGAGCGCTTC | 1870 |
| Qy | 612 | ACTTACAGAGTCGCGCCGAGAACCCGCTGGCGGCTGCCAGCGCCGCGCGCGCTAGTG | 671 |
| Db | 1871 | GTCCGGCAGGCGACCGGCAACGACGAGCGCGCGCGCCAAACCGACGTGTGAGCCTG | 1930 |
| Qy | 672 | GTGGACTCGGACCGCGGGACACGGCCAGCGCGCCCGGGACTTCCACGGCGCGCTCCCTG | 731 |
| Db | 1931 | ACTTGCCCGGTGCGCGCGGTGTAATGCGCGGCCCGCGGACAGCGGCGACGCCCTGCTG | 1990 |
| Qy | 732 | GCGCACCTGCAGCGGCGCGCGAGGCTATGCGGCGAGGCGGCCCGGCTCACGCGCC | 791 |
| Db | 1991 | GAGCGCAACTATCCCACTGGCGCGGAGTTCTCGGCGACGCGCGGACGTCAGCTTCAGC | 2050 |
| Qy | 792 | AGCACCGGCAGCGCACCTGCAGCGGTGACTGAAGCGCAAGCACGCGCGCCCTCAGCTGCTAC | 851 |
| Db | 2051 | ACCGCGCGCACGCGAAACTGGACGCTGGAGGGCTGCTCCAGCGGCACCGCGCACTGGAG | 2110 |
| Qy | 852 | GTGACCGGCGA | 862 |
| Db | 2111 | GAGCGCGGCTA | 2121 |

Search completed: September 19, 2004, 22:12:33
Job time : 106.092 secs

This Page Blank (uspto)


```

QY 61 CGCCACCGTATGATCAGACACAGTTACGGGGGCGCCCGCTTTCTCACCCGGGCCA 120
Db 61 CGCCACCGTATGATCAGACACAGTTACGGGGGCGCCCGCTTTCTCACCCGGGCCA 120
QY 121 AGGCTTTCGTGTCGTGGGCAAGGACGCCACCTCAGCTGCAGATCGTGGTAAATC 180
Db 121 AGGCTTTCGTGTCGTGGGCAAGGACGCCACCTCAGCTGCAGATCGTGGTAAATC 180
QY 181 CCAGCCACAGTGTGCTGGAGAGGACACAGCAGCGGTACCGCGCGCGCGCTTCC 240
Db 181 CCAGCCACAGTGTGCTGGAGAGGACACAGCAGCGGTACCGCGCGCGCGCTTCC 240
QY 241 GTCTGGCCAGGACCGCGACCTCTACCGCTCTACCTATCTCTGACCTGGCGCTGGCGACA 300
Db 241 GTCTGGCCAGGACCGCGACCTCTACCGCTCTACCTATCTCTGACCTGGCGCTGGCGACA 300
QY 301 GTGGCAATAGTGTGCGCGCGCGCAATGCAATAGCGAGGCGCTTCGTCGCGTGGGCC 360
Db 301 GTGGCAATAGTGTGCGCGCGCGCAATGCAATAGCGAGGCGCTTCGTCGCGTGGGCC 360
QY 361 TGCAGGTGGAACCGGAGGCGCGCTGCGCGAGCAGGCGCGCACTTCTGCTGGCGGCCA 420
Db 361 TGCAGGTGGAACCGGAGGCGCGCTGCGCGAGCAGGCGCGCACTTCTGCTGGCGGCCA 420
QY 421 CGTCCATCCGCGTGGCGAGGCTCAGAGGCACTTCCGCTGCGGTGGGTGGCTTCC 480
Db 421 CGTCCATCCGCGTGGCGAGGCTCAGAGGCACTTCCGCTGCGGTGGGTGGCTTCC 480
QY 481 CGAGCGCGGAGTGTGCTCAAGGACGCGCGCGCTGGGTGAGCGAGCGCGCC 540
Db 481 CGAGCGCGGAGTGTGCTCAAGGACGCGCGCGCTGGGTGAGCGAGCGCGCC 540
QY 541 CGGTGCGCGTGGAGAGCTCGCGAGGCAAGTGCGCTGCGCACTTCGCGCGCGCGCGC 600
Db 541 CGGTGCGCGTGGAGAGCTCGCGAGGCAAGTGCGCTGCGCACTTCGCGCGCGCGCGC 600
QY 601 CGGAGCGGCGACTTACAGGTGCGCGCGAGACCGCGCTGGGCGCTGCCAGCGCGCG 660
Db 601 CGGAGCGGCGACTTACAGGTGCGCGCGAGACCGCGCTGGGCGCTGCCAGCGCGCG 660
QY 661 CGGCGTAGTGTGACTCGACCGCGGACACCGGCGAGCGCGCGCGCGCGCGCG 720
Db 661 CGGCGTAGTGTGACTCGACCGCGGACACCGGCGAGCGCGCGCGCGCGCGCG 720
QY 721 CCGGCTCTTGGCGCACTTGCAGCGCGCGCGAGGCTATGCGCGCGAGGCGCGCGCG 780
Db 721 CCGGCTCTTGGCGCACTTGCAGCGCGCGCGAGGCTATGCGCGCGAGGCGCGCGCG 780
QY 781 CCTCACCGCCAGACCGGACCGGACCTGCGACCGGTGACTGAAGGCAAGCAGCGCGCC 840
Db 781 CCTCACCGCCAGACCGGACCGGACCTGCGACCGGTGACTGAAGGCAAGCAGCGCGCC 840
QY 841 TCAGCTGCTACGTGACCGGAGCGCCAGCGCGGACCGGTGTGGAAGGAAGGACGGCGAG 900
Db 841 TCAGCTGCTACGTGACCGGAGCGCCAGCGCGGACCGGTGTGGAAGGAAGGACGGCGAG 900
QY 901 TGGTGAACGAGGCGCGCGACCTGTGTGTACGAGGACGCGCAGGAGAACTTCGTCTCA 960
Db 901 TGGTGAACGAGGCGCGCGACCTGTGTGTGTACGAGGACGCGCAGGAGAACTTCGTCTCA 960
QY 961 AGATCTCTTCTGCAAGCAGTTCGACCGCGCGCTCTACCTGCAAGGCGTCAACCTCG 1020
Db 961 AGATCTCTTCTGCAAGCAGTTCGACCGCGCGCTCTACCTGCAAGGCGTCAACCTCG 1020
QY 1021 TGGGCCAGACCTAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1070
Db 1021 TGGGCCAGACCTAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1070

```

RESULT 2

US-10-077-130-6

; Sequence 6, Application US/10077130

; Publication No. US20020168742A1

GENERAL INFORMATION:

```

; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE REFERENCE: MPI2001-047PIRCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-077-130-6

```

```

Query Match          93.4%; Score 999; DB 14; Length 23907;
Best Local Similarity 100.0%; Pred. No. 1.7e-220;
Matches 999; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 ATGGATCAGCCACAGTTACGGGGGCGCCCGCTTTCTCACCCGGGCCAAGGCGCTTCGTG 131
Db 1 ATGGATCAGCCACAGTTACGGGGGCGCCCGCTTTCTCACCCGGGCCAAGGCGCTTCGTG 60
QY 132 GTGTGGTGGGCAAGGACGCCACCTCAGCTGCCAGATCGTGGGTAAATCCACGCCACAG 191
Db 61 GTGTGGTGGGCAAGGACGCCACCTCAGCTGCCAGATCGTGGGTAAATCCACGCCACAG 120
QY 192 GTGAGCTGGAGAGGACACAGCGCGGTGAGCGCGCGCGCGCTTCGTTCTGTCGCCCGAG 251
Db 121 GTGAGCTGGAGAGGACACAGCGCGGTGAGCGCGCGCGCGCTTCGTTCTGTCGCCCGAG 180
QY 252 GACGCGGACCTCTACCGCTCACTATCTGACCTGCGGCTGGGCGTGGGCGACAGTGGGCAATAC 311
Db 181 GACGCGGACCTCTACCGCTCACTATCTGACCTGCGGCTGGGCGTGGGCGACAGTGGGCAATAC 240
QY 312 GTGTCCGCGCGCGCAATGCCATAGGAGCGCTTCGTCGCGTGGGCGCTCAGGTGGAC 371
Db 241 GTGTCCGCGCGCGCAATGCCATAGGAGCGCTTCGTCGCGTGGGCGCTCAGGTGGAC 300
QY 372 GCGGAGCGCGGTGCGCGAGCAGCGCGCACTTCCTGCTGCGCGCCACAGTCCATCGC 431
Db 301 GCGGAGCGCGGTGCGCGAGCAGCGCGCACTTCCTGCTGCGCGCCACAGTCCATCGC 360
QY 432 GTGCGCGAGGGCTCAGAGGCGCACCTTCCGCTGCGCGCTGGGTGGCTTCCCGAGGCGCGCA 491
Db 361 GTGCGCGAGGGCTCAGAGGCGCACCTTCCGCTGCGCGCTGGGTGGCTTCCCGAGGCGCGCA 420
QY 492 GTGAGCTGGTCCAGGACGCGCGCGCTGGGTGAGCCCGAGCGCGCGCGCGCGGTG 551
Db 421 GTGAGCTGGTCCAGGACGCGCGCGCTGGGTGAGCCCGAGCGCGCGCGCGCGGTG 480
QY 552 GAGGAGCTCGCGAGGCAAGTGCCTGCGCAATTCGCGCGCGCGCGCGCGCGCGCGCG 611
Db 481 GAGGAGCTCGCGAGGCAAGTGCCTGCGCAATTCGCGCGCGCGCGCGCGCGCGCGCG 540
QY 612 ACTTACAGAGTCCGCGCGCGAGAAACCGCTGGGCGCTGCGAGCGCGCGCGCGCGCTAGTG 671
Db 541 ACTTACAGAGTCCGCGCGCGAGAAACCGCTGGGCGCTGCGAGCGCGCGCGCGCGCTAGTG 600
QY 672 GTGAGCTCGAGCGCGCGAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTAGTG 731
Db 601 GTGAGCTCGAGCGCGCGAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTAGTG 660
QY 732 GCGCACTTCAGCGCGCGCGAGAGGCTATGCGCGCGCGAGGCGCGCGCGCGCGCTCTAGTG 791
Db 661 GCGCACTTCAGCGCGCGCGAGAGGCTATGCGCGCGCGAGGCGCGCGCGCGCGCTCTAGTG 720
QY 792 AGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTAGTG 851
Db 721 AGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTAGTG 780

```


QY 852 GTGACCGCGGAGCCAGCCGAGACGGTGTGGAAGAGGACGCCAGCTGGTGACCGAG 911
Db 781 GTGACCGCGGAGCCAGCCGAGACGGTGTGGAAGAGGACGCCAGCTGGTGACCGAG 840
QY 912 GCGCGGCGGACGCTGGTGTACGAGGACGCGCAGGAGAACTTCGTCTCAAGATCTCTTC 971
Db 841 GCGCGGCGGACGCTGGTGTACGAGGACGCGCAGGAGAACTTCGTCTCAAGATCTCTTC 900
QY 972 TGAAGCAGTCCGACCGCGGCTCTACACCTTGACGCGGCTCCAACTCTGTGGGCCAGACC 1031
Db 901 TGAAGCAGTCCGACCGCGGCTCTACACCTTGACGCGGCTCCAACTCTGTGGGCCAGACC 960
QY 1032 TACAGCTCTGTGCTGTGTAGTCCGCGGAGCCCGCGTT 1070
Db 961 TACAGCTCTGTGCTGTGTAGTCCGCGGAGCCCGCGTT 999

RESULT 3
US-10-094-1001/c
; Sequence 1001, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR FILING DATE: 2002-01-24
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1001
; LENGTH: 2884
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-1001

Query Match 23.8%; Score 255; DB 16; Length 2884;
Best Local Similarity 100.0%; Pred. No. 1.6e-49;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 805 GCACCTGCACGGTGTCTGAAGCAGGACCGCGGCTCAGCTGCTAGTACCGGCGAGC 864
Db 634 GCACCTGCACGGTGTCTGAAGCAGGACCGCGGCTCAGCTGCTAGTACCGGCGAGC 575
QY 865 CCAAGCCCGAGACGGTGTGGAAGAGGACGCGCAGCTGGTACCGAGGCGCGGCCACG 924
Db 574 CCAAGCCCGAGACGGTGTGGAAGAGGACGCGCAGCTGGTACCGAGGCGCGGCCACG 515
QY 925 TGGTGTACGAGACCGCAGGAGAACTTCGTCTCAAGATCCTTTTCGAAGCAGTCGG 984
Db 514 TGGTGTACGAGACCGCAGGAGAACTTCGTCTCAAGATCCTTTTCGAAGCAGTCGG 455

QY 985 ACCGCGGCTCTACACCTTGACGCGGCTCAACCTCTGTGGCCAGACCTACAGCTCTGTGC 1044
Db 454 ACCGCGGCTCTACACCTTGACGCGGCTCAACCTCTGTGGCCAGACCTACAGCTCTGTGC 395
QY 1045 TGGTGTAGTGGCG 1059
Db 394 TGGTGTAGTGGCG 380

RESULT 4
US-09-764-891-6043
; Sequence 6043, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6043
; LENGTH: 9591
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6043

Query Match 18.5%; Score 197.6; DB 10; Length 9591;
Best Local Similarity 51.7%; Pred. No. 2e-36;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCAGCAGCCACACTCCGGCCCTGCCCCAGAGCCCCCATAGAGAGGTCCTCCCGCCCA 66
Db 2 CCTCCGACCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 61
QY 67 CGGTATGATCAGCACAAGTTACGCGGCGCCCGCTTTCTTCTACCCGCGCCCAAGGCT 126
Db 62 TGAAGCGCAGCTCGGGGATCAGGGGAGCCCCCTGCTTCTCGCTTCCCGCGGCTG 121
QY 127 TCGTGTGTGGTGGCAAGGACCCCTCAGCTGCCAGATCGTGGTAAATCCACGCG 186
Db 122 TCGGGTGTAAAGTGGCGCCGAGGCGAGCTCAAGTGTGTCTTCTGGGGAGCCGCC 181
QY 187 CACAGGTGAGCTGGGAGAGGACCAAGCCGCTGACCGCGCGCGCTTCCTCGTCTGG 246
Db 182 CTGTAGTGTGGGAGAGGCGGCGCAGCAGCTGGCGGCTCGGAACCTGAGCTTCC 241
QY 247 CCGAGCAGCGGACCTCTACCGCTCACTATCTCTGACCTGGCGTGGGGGACAGTGGC 306
Db 242 CCGCGCAGCGCGGAGCAGCGGCTCTGCTGACCGCCGCACTGCCACCGCGGGG 301
QY 307 AATACGTGTCCGCGCGCAATGCCATAGGCGAGGCTTCGCTGCGTGGGCTGCGAG 366
Db 302 TCTACGTGTCCGCGCGCAACCGCGCGCGGAGGCTTACGGGCGCGCGCTCACCG 361
QY 367 TGAACCGGAGCGCGCTGCGCGGAGCAGCGCGGCACTTCTCTGTCGCGCCACGTTCA 426
Db 362 TGTGAGCGCGCGCTCCGACCCCGAGCTCAGCCCGCGGAGCGCGCTGTCATCGC 421
QY 427 TCCGCGTGGCGGCTCAGAGGCGCCTTCCTCGTGGCG-----CGTGGGTGGCTCC 480
Db 422 CGGGTCCGGAGGCGCGCGCTTCTCTACGGGCTCGATCCAGTGGGTGCTGC 481
QY 481 CGAGGCGCGGAGTGTGTCTCAAGGACGGCGCGCTGGGTGAGCCCGAGCGGCCCG 540
Db 482 GGGGGCGGAGTGTGTGAGCTGCGCGGCGGGGCGCTCCCGAGCCACACTGTACT 541
QY 541 GCGTGGCGTGGAGGAGCTCGGCGAGGCAAGTGGCTGCGCATTCGGGCGCGCGCGCG 600
Db 542 GGGAGAAGCAGGATGGCTTGGAGAAAGTGTGGGACAGCAGCCACTTCGGGCTCAGC 601
QY 601 GCGACGCGCGCACTTACGAGTCCGCGCGAGAACCCGCTGGGC--GCTGCCAGCGCCG 658

Db 602 CGGGCGCGCGAGGACGGCCCGCGCGAGCGCTGGCACTGCGCATCTTGGCGCTCGGC 661
QY 659 CGCGCGCTAGTG-GTGAGACTGAGACGGCGCGGGAACGGCCAGCGCGCGGACCTTCCA 717
Db 662 TGCGGATTCGGCGCTCTACGTGTGCCACGCGCCGCAACGGCGACGGCCACGCGAGCGG 721
QY 718 CGCGCGCGCTCTTGGCGCACTGACGCGCGCGCGCGAGGCTATCGCGCGCGAGCGCGCC 777
Db 722 GGGCGCTGCTCAGGTGACACAGCCCCCGGAGAGCGCCCGCGGACCCGACAGGCC 781
QY 778 CGCGCTC-----ACGCGCCAGCACCGGCAACGGCACTGCGCACTGCGTGAAGGCA 828
Db 782 CGCGCGCGGTGGAGCGCTCAAGTGGCGCGCTTAAGACCTTCTGGGTGAACGAGGCA 841
QY 829 AGCAGCGCGCTCAGCTGCTACGTGACCGCGCGGAGCCCAAGCGCGAGACGCGTGTGAAGA 888
Db 842 AGCAGCGCAAGTTCGCTGCTACGTGATGGCAAGCGCGGAGCTTCTGGGTGAACGAGGCA 901
QY 889 AGCAGCGCGCACTGCTGACCGAGCGCGCGCGCGCGCGCGCTTACGAGACGCGCGAGG 948
Db 902 GGGAGGCGCGCGCTGCTCCCGGACCGCGCGCGCTTACGAGACGCGCGAGCGCG 961
QY 949 ACTTCGTGCTCAAGTCTCTTTCGCAAGCAGTGGAGCGCGCGCGCTTACCTGACGCG 1008
Db 962 GCTTCGTGCTCAAGTCTCTTTCGCAAGCAGTGGAGCGCGCGCGCTTACCTGACGCG 1021
QY 1009 CGTCAACCTCGTGGCGCAGACCTACAGCTCTGTGCTGGTGTAGTGGCGGAGCGCGCG 1068
Db 1022 CGCGCACTCGCGCGCGCAGCGCTCAGTGGCGTGCAGCTGCACGTTGAAGGTACGCGG 1081

RESULT 5
US-10-091-438-268
; Sequence 268, Application US/10091438
; Publication No. US20030077606A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT217C1
; CURRENT APPLICATION NUMBER: US/10/091.438
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,879
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270

; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937

```
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
```

```
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match      18.5%; Score 197.6; DB 15; Length 9591;
Best Local Similarity 51.7%; Pred. No. 2e-36;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY      7 CCAGCAGCCACACTCCGGCCGCTGCCAGAGCCCCCATAGAGAGAGTCCCGCGGCCA 66
Db      2 CCCTCGCACCCCCCTGCCCCCACCCTTCCGCGCTGCAGCGGTCCGCGCGCGCA 61
QY     67 CGGTATGATCAGCCACAGTTTCAGGGGGGCGCGCTTTCTCACCAGGCAAGGCT 126
Db     62 TGAAGCGGAGCTCGGGGATCAGGAGAGCCCCCTGCTTCTCGCTTCCCGCGCTG 121
QY    127 TCGTGTGTGCTGGGCAAGGACGCCCTCAGCTGCCAGATCGTGGTAAATCCACACG 186
Db    122 TCGGGTGTAGTGGCGCGAGGCGAGCTCAAGTGGTGTCTCTGGGGAGCGCGC 181
QY    187 CACAGTGAGCTGGGAGAGGACACAGAGCCGGTGAAGCGCGCGCGCTTCGCTCTGG 246
Db    182 CTGTAGTGTGGGAGAGGCGGCGAGCTGCGGCTCGGAACCCCTGAGCTTCC 241
QY    247 CCCAGCAGCGACCTTACCGCTCAGCTATCTGACCTGCGCTGGCGCAGTGGCG 306
Db    242 CGCGGAGCGCGCGGAGGAGCGGCTGCTGCTGACCGCCCATGCCACGAGCGCGGG 301
QY    307 AATACGTGTCCGCGCGCAATGCAATAGGCGAGGCTTCTGCTGCGTGGGCTGACAG 366
Db    302 TCTACGTGTCCGCGCGCAACGCGGCGCGGAGGCTTACGCGGCGCGCGCTCACCG 361
QY    367 TGAACCGGAGCGCGGTGCGGAGAGCGCGGCTTCTGCTGCGGCGCCACGTTCA 426
Db    362 TGTGAGCGCGCGGCTCCGACCCCGAGCTGACGCCCGCGCGCGCTGCGCATCGC 421
QY    427 TCCGCGTGGCGAGGCTCAGAGGCGACCTTCCGCTGCGG-----CGTGGGTGGCTCC 480
Db    422 CGGGGTCCGGGAGGCGCGCGCTTCTTCTCAGGGGCTCGATCCAGTGGGTGCTGC 481
QY    481 CGAGCGCGCAGTGAAGTGTTCAGGAGCGGGCGGCGCTGGGTGAGCGCGCGCGCC 540
Db    482 GGGGCGCGAGGTGTGTGCTGAGCTGCGGGCGGGGGCTTCCCGAGCCACACTGTACT 541
QY    541 GGTGCGCTGGAGGAGCTCGGCGAGGCAAGTGCCTGCGCATTCGGGCGCGCGCGCG 600
Db    542 GGGAGAGGAGCGGATGCGCTTGGACAGAGTGTGGAGCAGCAGCAGCTTCCGCTCCAG 601
QY    601 GCGACCGCGGCACTTACGAGGTCCGCGCCAGAAACCGCTGGGC--GTTGCCAGCGCGC 658
Db    602 CGGGCGCGCGAGGAGCGCGCGCGGCGAGCTGGCACTGGCGCATCTTGGGGGTTCGCG 661
QY    659 CGCGGCGCTAGTG--GTGGACTCGGAGCGCGGAGCAGCGGCGCGCGCGCGCGCTCCA 717
Db    662 TGCCGATTCGGGCGTCTACGTGTGCCACGCCCGCAACCGCAGCGGCGCGCGCGCG 721
QY    718 CGGCGCGCTTCTTGGCGCACCTGACGCGCGCGCGGAGGCTATGCGCGCGCGCGCGCG 777
Db    722 GGGCGCTGTCTCAGGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
QY    778 CCGCTC-----ACCGCCAGCAGCGCGCAGCGCGCGCGCGCGCGCTGACGTGAAGGCA 828
```


;
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038

;
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08

```
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match      18.5%; Score 197.6; DB 15; Length 12415;
Best Local Similarity 51.7%; Pred. No. 1.9e-36;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCAGACGCCACATCTCGGCGGTGCCCCAGAGCCCAATAGAGAGAGTCCCGCGCGCA 66
Db 2 CCCTCCGACACCCCGCTGCCCGCCACCGTTCCGCGCTGCAGCGCGTCCGCGCGCA 61
QY 67 CCGTCATGGATCAGCCACAGTTAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 126
Db 62 TGAAGGCGAGCTCGGGGGGATCAGGGGAGCCCGCGCTTCTCCGCGCTTCCCGGGCTG 121
QY 127 TCGTGGTGTTCGTTGGGCAAGGAGCCACCTTCAGCTGCCAGATCGTGGGTAAATCCACGC 186
Db 122 TCGGGGTGGTAAAGTGGCGCGAGCCGAGCTCAAGTGCCTGGTGGTCTTGGGGAGCCCGCG 181
QY 187 CACAGGTGAGCTGGGAGAGGACACGAGCCGCTGAGCGCGCGCGCTTCGCTTGG 246
Db 182 CTGTAGTGGTGGGAGAGGGGGGCGAGCAGCTGGCGGCTTCGAAACGCTTCCAGCTTCC 241
```

```
Db 542 GGGAGAGGACGGGATGCGCCCTGGAGGAAGTGTGGACAGAGCCACTTCCGGCTCCAGC 601
QY 601 GCGAGCGCGGCACTTACGAGGTCCGCGCGGAGAACCCCGCTGGGC--GCTGCCAGCGCGC 658
Db 602 CGGGCGCGCGGAGACGGCGCCCGGCGAGCTTGGCACTCCGCGATCCTGGCGGCTCGGC 661
QY 659 CGCGCGGTAGTG-CTGGACTCGGACGCGCGGAGACAGCGCCAGCGCGCCGGACCTTCA 717
Db 662 TGCCCGGATTCCGGCGCTTACGTGTGCCACGCGCCCGCAACGCGCACGGGCCACGCGAGCGG 721
QY 718 CGGCGCGCTCTCTGGCGCACCTGTCAGCGCGCGCGCGAGGCTATGCGCGCCGAGGCGCGCC 777
Db 722 GGGCGCTGCTCCAGGTGCACAGCCCGCGAGAGCCCGCGCGGACCCCGACGAGGCGCC 781
QY 778 CGGCCTC-----ACGCGCCAGCACCGGACCGGACCTGACGCTGACTGAAGGCA 828
Db 782 CGCGCGCGGTGTGTGAGCGCTCAAGTGCAGCGCTTAAGACCTTCTGGGTGAACGAGGCA 841
QY 829 AGCAGCGCGCTCAGCTGCTACGTGACCGCGAGCCCAAGCCGAGACGCTGTGGAAGA 888
Db 842 AGCAGCGCAAGTTCCGCTGCTACGTGATGGGAAAGCCCGAGCCCGAGATCGAATGGCACT 901
QY 889 AGGACGGCCAGCTGGTGACCGAGCGCGCGCCACGTTGTACGAGACCGCAGGAGA 948
Db 902 GGGAGGGCGCGCGCTGCTCCGAGACCGCGCGCGCTCATGTACCGGACCGCGAGCGCG 961
QY 949 ACTTCGTGCTCAAGATCCTTCTGCAAGCAGTGGAGCGCGCGCGCTTACACCTTGCACCG 1008
Db 962 GCTTCGTGCTCAAGTGTCTTACTGCGAGCGCAAGGATCGTGGGCTTCTGCTGCGCGC 1021
QY 1009 GGTCAACCTCGTGGCGGACACCTACAGCTCTGTGCTGTGTAGTGGCGAGCGCGCG 1068
Db 1022 CGGCAACTCGCGCGGCGAGACGCTCAGTGCCTGCGAGCTGCAGTGAAGGTACGCGCG 1081

RESULT 8
US-09-764-891-6045
; Sequence 6045, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6045
; LENGTH: 20565
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6045

Query Match      18.5%; Score 197.6; DB 10; Length 20565;
Best Local Similarity 51.7%; Pred. No. 1.8e-36;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCAGACGCCACATCTCGGCGGTGCCCCAGAGCCCAATAGAGAGAGTCCCGCGCGCA 66
Db 2 CCCTCCGACACCCCGCTGCCCGCCACCGTTCCGCGCTGCAGCGCGTCCGCGCGCA 61
QY 67 CCGTCATGGATCAGCCACAGTTAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 126
Db 62 TGAAGGCGAGCTCGGGGGGATCAGGGGAGCCCGCGCTTCTCCGCGCTTCCCGGGCTG 121
QY 127 TCGTGGTGTTCGTTGGGCAAGGAGCCACCTTCAGCTGCCAGATCGTGGGTAAATCCACGC 186
Db 122 TCGGGGTGGTAAAGTGGCGCGAGCCGAGCTCAAGTGCCTGGTGGTCTTGGGGAGCCCGCG 181
QY 187 CACAGGTGAGCTGGGAGAGGACACGAGCCGCTGAGCGCGCGCGCTTCGCTTGG 246
Db 182 CTGTAGTGGTGGGAGAGGGGGGCGAGCAGCTGGCGGCTTCGAAACGCTTCCAGCTTCC 241
```

QY 247 CCCAGGCGGACCTTACGGCTCACTATCTTGGACCTTGGCGTGGGCGACAGTGGCG 306
Db |||||
QY 242 CGGCGGACGGCGGACGACGGCTGTGTGTACCGCGGACTGCCACCGACGCGGGGG 301
Db |||||
QY 307 AATACGTGTGCGCGCGCGCAATGCCATAGCGGAGGCTTTCGTGCGTGGGCTGACGG 366
Db |||||
QY 302 TCTACGTGTGCGCGCGCGCAACGCGCGCGGCGGAGGCTTACGCGCGCGCGCGCTCACCG 361
Db |||||
QY 367 TGGACGCGGAGGCGCGTGTGCGCGAGGCGGCGCATCTTCTGCTGCGCGCCACGTCGA 426
Db |||||
QY 362 TGCTGGAGCGCGCGGCTTCGACCGCGAGTGTGACCGCGCGCGCGCTGCTGCTGCTG 421
Db |||||
QY 427 TCCGCGTGTGCGGAGGCTTCAGAGCGCCACCTTTCGCTGCGG-----CGTGGGTGGTCCC 480
Db |||||
QY 422 CGGCGTCCGCGGAGGCGCGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 481
Db |||||
QY 481 CGAGCGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db |||||
QY 482 GGGGCGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541
Db |||||
QY 541 GCGTGTGCGGAGGCTTGGCGGAGGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db |||||
QY 542 GGGAGAGGAGCGGATGCGCTTGGAGAGTGTGGGACAGCAGCCTTTCGCGCTCCAGC 601
Db |||||
QY 601 GCGAGCGGCGCACTTACAGAGTTCGCGCGGAGAACCGCGTGGGC--GCTGCGAGCGCGC 658
Db |||||
QY 602 CGGCGCGCGGAGGAGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 661
Db |||||
QY 659 CGCGCGGCTAGTG--GTGAGCTCGGAGCGCGCGGAGCAGCGCGCGCGCGCGCGCGCTCCA 717
Db |||||
QY 662 TGCGGAGTTCGCGGCTGTACGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 721
Db |||||
QY 718 CGGCGCGCTCTGCGCGCACTGAGCGCGCGCGCGGAGGCTTTCGCGCGGAGGCGCGCGC 777
Db |||||
QY 722 GGGCGCTCTCCAGGTGCAACAGCGCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGC 781
Db |||||
QY 778 CGGCTC-----ACGCGCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 828
Db |||||
QY 782 CGCGCGCGGTGTGAGCGCGCTCAAGTGTGCGCGCTTACGCTTCTGCGTGAACGAGGCA 841
Db |||||
QY 829 AGCAGCGCGCTCAGCTGTCTACGTGACCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 888
Db |||||
QY 842 AGCAGCGCAAGTTCGCTGTCTACGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 901
Db |||||
QY 889 AGGACGCGGAGTGTGTACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948
Db |||||
QY 902 GGGAGGCGCGCGCGCTGTCTCCGCGACCGCGCGCGCTCATGTACCGGACCGCGCGCGCG 961
Db |||||
QY 949 ACTTGTGTCTCAAGATCTTCTTCTCAAGCAGTGTGCGCGCGCGCGCGCTTACACCTGCA 1008
Db |||||
QY 962 GTTGTGTCTCAAGTGTCTTACTCCAGGCGCAAGGATCTGCGCTTACGCTCTGCGCGC 1021
Db |||||
QY 1009 CGTCAACCTCGTGGCGCGCAGCTACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1058
Db |||||
QY 1022 CGCGCAACTCGCGCGCGCGCGCGCTCAGTGTGCGCGCGCGCGCGCTGCAAGGATGCGG 1081
Db |||||

RESULT 9

US-10-091-438-270/c
; Sequence 270, Application US/10091438
; Publication No. US20030077606A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ17C1
; CURRENT APPLICATION NUMBER: US/10/091,438
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,879
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628

; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01

[illegible]

QY 427 TCCGCTGCGGAGGGCTCAGAGCCACCTTCCTGCTGCCG-----CGTGGGTGGCTCC 480
Db 20144 CGGGGTCCGGGAGGGCGCCCGCTCTTCCTCAGGGGCTCGATCCAGTGGGTGCTGC 20085
QY 481 CGAGCGCGGAGTGTGCTGCTCAAGAGAGCGGGCGGCTTGGGTGAGCCCGACGCGCCCG 540
Db 20084 GGGGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 20025
QY 541 GCCTGCGCGTGGAGAGCTCGGCGAGGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 20024 GGGAGAGGACGGGATGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19965
QY 601 GCGAGCGCGGCTTACGAGTCCG 658
Db 19864 CGGGCGCGCGCGGAGGACCG 19905
QY 659 CGCGCGCTAGTG--GTGAGCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
Db 19904 TGCGGATTCGGGCTCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19845
QY 718 CGGGCGCGCTTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777
Db 19844 GGGGCTGCTCTGAGTGCAGGACCG 19785
QY 778 CCGGCTC-----ACCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828
Db 19784 CCGGCGCGTGTGGAGCGCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19725
QY 829 AGCAGCGCGCTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 888
Db 19724 AGCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19665
QY 889 AGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948
Db 19664 GGGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19605
QY 949 ACTTGTGCTCAAGATCTCTTCTCAAGAGTCCGACCGCGCTTCAACCTGCAAGTCCGCGCG 1008
Db 19604 GTTGTGCTCAAGTCTTCTTCTCAAGAGTCCGACCGCGCTTCAACCTGCAAGTCCGCGCG 19545
QY 1009 GTCCAACTGTGGCGAGACTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1068
Db 19544 CGCGCACTCGCGCGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19485

RESULT 10
US-10-040-739-716
; Sequence 716, Application US/10040739
; Publication No. US20020173635A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John
; Lavallie, Edward
; Racie, Lisa
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
; NUMBER OF SEQUENCES: 1519
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/040,739
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/036,520
FILING DATE: 03-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELEPHONE: (617) 498-8224
TELEFAX: (617) 878-5851
INFORMATION FOR SEQ ID NO: 716:
SEQUENCE CHARACTERISTICS:
LENGTH: 471 base pairs
TYPE: nucleic acid
STRADEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 716:
US-10-040-739-716
Query Match 11.0%; Score 117.2; DB 14; Length 471;
Best Local Similarity 65.9%; Pred. No. 8.8e-18;
Matches 170; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 807 ACTGCACGCTGACTGAAGCAAGCAGCGCGCCCTCAGCTGCTAGTGCAGCGCGCGCGCGCGCG 866
Db 23 ACCTTCTGGTGAACGAGGCGCAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82
QY 867 AGCCCGGACGCTGTGGAGAGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 926
Db 83 GAGCCCGAGATCGATGGCAGCTGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142
QY 927 GTGTACGAGGACCGCGCAGGAGAACTTCGTGCTCAAGATCCTCTTGTGCAAGAGTCCGAC 986
Db 143 ATGTACCGGACCGCGCAGCGCGCTTCGTGCTCAAGTCTCTTACTGCCAGGCCAAGGAT 202
QY 987 CGCGGCTCTACACCTGACGCGCGCTCAACCTGCTGGGCGACAGCTTACAGCTCTGTGCTG 1046
Db 203 CGTGGGCTCTACGCTGCTGCG 262
QY 1047 GTCGTAGTGGCGAGCGCC 1064
Db 263 CTCACGCTGAAGAGCGCC 280

RESULT 11
US-09-764-853-138
; Sequence 138, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: FJZ06
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 138
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-138
Query Match 8.1%; Score 86.6; DB 9; Length 609;
Best Local Similarity 51.3%; Pred. No. 8.9e-11;
Matches 299; Conservative 0; Mismatches 254; Indels 30; Gaps 3;
QY 377 GGCGCGTGGCGGAGCAGCG 436
Db 57 GCCTGGGTCCGGGAGGCG 116

QY 437 CGAGGGCTCAGAGGCACTTCCGGTGGCGGTGGGTGGCTCCCGAGGCGCGGAGTGAG 496
 Db 117 GCGGGGGGCGAGGGTGGTGTGACGTGCGGGCGGGGGGCGCTCCCGAGCCACACATGTA 176
 QY 497 CTGGTCCAAAGGACGGCGCGCTGGGTGAGCCCGACCGGCCCGCGTGGCGTGGAGGA 556
 Db 177 CTGGGAGAGAGAGGAGTGGCCCTGAGCAAGTGTGGAGAGAGACCACTTCGCGCTCCA 236
 QY 557 GCTCGGAGAGGCAAGTGGCTGCGCAATCGGGCGCGCGCGCGCGCGCGCGCACTTA 616
 Db 237 GCGGGGCC-----GCGCGAGAGACGCGCCCGCGCGGAGCTGGCAATTCGCATCTG 288
 QY 617 CGAGGTCCGCGCGAGAGACCGCGCTGGGCGCTGCCAGCGCGCGCGGGCGGTAGTGTGA 676
 Db 289 GCGGCTTCGGCTGCGGATTCGCGGTCTACGTGTGCCACGCGCGCGCGCGCGCGG----- 339
 QY 677 CTCGAGCCCGCGGACACGGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 736
 Db 340 -----CACGGCCACGCGAGCGGGGCGCTGCTCCAGGTGCAACGCGCGCGCGAG 390
 QY 737 CCGTCAAGCGCGCGCGGAGGTATGCGCGCGCGAGGGGCGCGCGCGCGCGCGCGCGCG 796
 Db 391 AGCG 446
 QY 797 CGGACG 856
 Db 447 CGCGCTTAAGACCTTCTGGGTGAACAGAGGGCAAGCAGCGCGCGCGCGCGCGCGCG 506
 QY 857 CGGCGAGCG 916
 Db 507 GGGCAAGCG 566
 QY 917 GCGGCACGTGTACAGGAGCGCGAGGAACTTCGTGCTC 959
 Db 567 CCGCGCGCTCATGTACCGGACCGCGAGCGGGCTTCGTGCTC 609

RESULT 12

US-10-425-114-26383
 ; Sequence 26383, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E.
 ; APPLICANT: Tabaska, Jack E.
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 26383
 ; LENGTH: 1431
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB4171-034-C11_FLI
 US-10-425-114-26383

Query Match 7.5%; Score 80.4; DB 13; Length 1431;
 Best Local Similarity 44.3%; Pred. No. 2.1e-09;
 Matches 423; Conservative 0; Mismatches 521; Indels 10; Gaps 2;
 QY 21 TCCGGCGCGTCCCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 79
 Db 49 TCCATCGCGCTTCTTGGACTCCATCGGAGCGCGCAAGGAGAACCAATGACGACGCGAC 108
 QY 80 GCACAGTTACGCGGGCGCGCGCGCTTCTCACCGCGCGCGCGCGCGCGCGCGCGCGCG 139

Db 109 GAAGAGCTCCGGCGCGTGCACACGCTCGGCGCGCGCGCGCGCGCGCGCGCGTGTGTGCT 168
 QY 140 GGCAAGAGACGCCACCTCAGCTGCCAGATCGTGGTAAATCCACGCCACACAGTGTGCTG 199
 Db 169 GGGTCCGAGAGGCTCGGGGAGGTGGTGGCGGTCAAGTCGGCGCGCGCGCGCGCGCGCG 228
 QY 200 GGAAAGGACCAAGAGCGCGGTGACCGCGCGCGCGCGCGCTTCGTCTGGGCCAGGACGGG 259
 Db 229 CGCGCGCAGCTCCAGCGCGAGGCGCGCTCTCCGCGCGCGCTCTCGTCCGCGCACATCGT 288
 QY 260 CTTTACCGCGCTCATCTTGCAGCTCGCGCTGGCGGTGGCGACAGTGGGCAATAGTGTGCG 319
 Db 289 GCGCTGCTCGGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGCAT 348
 QY 320 CGCGCGCAATGCATAGCGAGGCGCTTCCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 379
 Db 349 GCGCGCGCGGTGCTCG 408
 QY 380 CGCGTGGCGCGAGAGCG 439
 Db 409 CCGCATCGCGCGCTACGCGCGGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 468
 QY 440 GGGCTCAGAGGCGCACTTCCGCTGCGG-----CGTGGGTGGCTCCCGAGGCGCGCG 490
 Db 469 GCTGCTGACGGGAGCTCAAGGCCCGGAACTGCTGCTCATCGCGCGCGCGCGCGCGCGCG 528
 QY 491 AGTGAAGTGGTCCAGAGAGCG 550
 Db 529 GCTGACCGACTTCGGGTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 588
 QY 551 GGAGAGTTCGCGCGAGGCAAGTGGCGCTCGCGCTTCGGCGCGCGCGCGCGCGCGCGCGCG 610
 Db 589 CCGCGCGTTCATGGCGCGCGAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648
 QY 611 CACTTACAGAGTTCGCGCGCGGAGAACCGCTGGCGCGCTGCCAGCGCGCGCGCGCGCGCT 670
 Db 649 CTGGCGCTCGGGTGCATGCTGCTGAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708
 QY 671 GGTGACTCGGACCGCGCGGACACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730
 Db 709 GGACGAGAGCGACCTCTCTCGCGCGCTCCACCGATCGGTACAGGACGACGTCGCGCG 768
 QY 731 GCGCACCTGACG 790
 Db 769 GGTGCGCGGTGGTGTGCGCGCGGAGCAAGACTTCTGCGCGCGCTGCTTCGAGCGCGCG 828
 QY 791 CAGCACCGGACCG 850
 Db 829 CCG 888
 QY 851 CGTGAACCGCGAGCG 910
 Db 889 CGCGCGCTGCG 948
 QY 911 GGGCG 964
 Db 949 CACGCTGCACGACCGCTTCTGGGACTCGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1002

RESULT 13

US-10-156-761-6483
 ; Sequence 6483, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262

```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6483
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2079)
; US-10-156-761-6483

Query Match      7.4%; Score 79.2; DB 15; Length 2079;
Best Local Similarity 47.7%; Pred. No. 3.8e-09;
Matches 358; Conservative 0; Mismatches 383; Indels 9; Gaps 4;

Qy 65 CACCGTCATGATCAGGCACAGTTCAGCGGGGGGCGCCGCTTCTACCCGCGCCCAAGGC 124
Db 672 CGCGGGCAGCCCTCGCCCTGCTCTCGCGGGAGCTGGCTGGCGCCGCCGCCCGCCAGGC 731

Qy 125 CTTCTGCTGCTGGGGAAGGACGACCCCTCAGCTGCAGATCGTGGGTAAATCCAC 184
Db 732 CGCGCTGCCGACTCGCCAAAGCAGCTGCGCTCCGAGGGCA---CGAGGGCCCGCGAC 788

Qy 185 GCCACAGGTGAGTGGGAGAAGGACAGCAGCCGCTGACGCGCGCGCGCTTCGCTCT 244
Db 789 GGCACGCTCTTCCGGCTCGCTACGCGGGGCTCGCGGCCACCGCCCGAGCGGATCTGCG 848

Qy 245 GGCCAGGACGGGACCTCTACCGCTCACTATCTGACCTGCGCTGGCGGACAGTGG 304
Db 849 CTTGCTCTCCCTCGCGCCCGCGGCGCTGCTGCGACCGGACACCGCTTCGCGCTCGTGG 908

Qy 305 GCAATACGTGTCGCGCGCGCAATGCCATAGCGGAGCCCTTCGCTCCGCTGGCGCTGCA 364
Db 909 CTGCTCGGTGAGCGGGGCCACAGACCTCGACGATTGCGCGCGCTCGCCCTCGTGGCG 968

Qy 365 GGTGACGCGGAGGCGGCTGCGCGGAGCAGGCGCGGCACTTCCTGCTGGCGCCAGTGC 424
Db 969 CTGCTGGAATCCCGCTGCGCACAGTACGAGGTGCGCGGCTTCCTGCTGGCGCTGGCG 1028

Qy 425 CATCCGCTG---CGCGAGGGCTCAGAGGCACTTCCTCGCTGCGCGGTGGGTGGCTCCC 480
Db 1029 GCGCTCGCGAGTACAGGAGCGTCCCGCGAGCTTCAGCTGCGCGCGCGCGGATGCT 1088

Qy 481 CGAGCCCGGCACTGAGCTGTGTCAGGACGAGGCAAGTGGCTGCGCTGCGCATTCGGCGCGCGCGCG 600
Db 1089 CGA-GCGGACCGTGGGCTGTGCACTCTGCGGGCCATCACCGAGACGAGCATCG 1147

Qy 541 GGTGCGGTGAGGAGCTCGCGAGGCAAGTGGCTGCGCTGCGCATTCGGCGCGCGCGCGCGCG 600
Db 1148 AGCGCGCGGAGAGCTCTCGGGATGCCCCGACGCTGCGCTTCCGNAACCCCGCGCGCGG 1207

Qy 601 GCGAGCGGCGCACTTACAGAGTCCGCGCGGAGAACCCGCTGGCGGTGCTCCAGCGCGCGCG 660
Db 1208 CCGCGGCTGCTGCGCGGTGCGGCGAGCGCGCGCTTGGCGCGCGCTCGAGCGCGGTGCG 1267

Qy 661 CCGCGCTAGTGTGACTCGGACGCGGAGACAGCGCGCGCGCGCGGAGCTCCAGG 720
Db 1268 CCGAGCGGAGTGGACACCTTGGCGCGCGGCTCATGTCCAGCTGTGCGCGGATGG 1327

Qy 721 CCGCGCTCT-GGCGCACTGCGAGGCGCGCGGAGGCTATGCGCGCGCGAGGCGCGCGCG 779
Db 1328 TGCCCACTTCGGCAGCGAGGCTGCGCGCGGAGCTGTACGGCATCCACCGCTGGTCC 1387

Qy 780 GCTCACCGCGCAGCACCGGCAACCGGCAACC 809
Db 1388 TCGAGCTCGCGGCGCGGAGCCTGCCCC 1417
```

```

RESULT 14
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
; US-10-156-761-1

Query Match      7.4%; Score 79.2; DB 15; Length 9025608;
Best Local Similarity 47.7%; Pred. No. 1.1e-09;
Matches 358; Conservative 0; Mismatches 383; Indels 9; Gaps 4;

Qy 65 CACCGTCATGATCAGGCACAGTTCAGCGGGGGGCGCCGCTTCTACCCGCGCCCAAGGC 124
Db 7794391 CGCGGGCAGCCCTCGCCCTGCTCTCGCGGGAGCTGGCTGGCGCCGCCGCCCGCCAGGC 7794450

Qy 125 CTTCTGCTGCTGGGGAAGGACGACCCCTCAGCTGCAGATCGTGGGTAAATCCCAAC 184
Db 7794451 CGCGCTGCGCGACCTCGCCAGCAGCTGCGCTCCGAGGGCGA---CGAGGGCCCGCGAC 7794507

Qy 185 GCCACAGGTGAGTGGGAGAAGGACAGCAGCGGCTGAGCGGGCGCGCGCTTCGCTCT 244
Db 7794508 GGCACGCTCTTCGCGCTCGCTACGCGGGTTCGCGGCAACCGCCAGCGGATCTGCG 7794567

Qy 245 GGCCAGGACGGGACCTCTACCGCTCACTATCTGAGACCTGGGCTGGGCGACAGTGG 304
Db 7794568 CTTGCTCTCTCGCGCGCGCGGCTCTGCGACCGGCAACCGCTCGCGCTCGCGCG 7794627

Qy 305 GCAATACGTGTCGCGCGCGCAATGCCATAGCGAGGCTTCGCTGCGCGTGGGCTGCA 364
Db 7794628 CTGCTCGGTGAGCGGGCCACACAGACCTTCGCGCGCTCGGCTCGGCTCGTGG 7794687

Qy 365 GGTGAGCGGAGGCGCGCTGCGCGGAGCGCGCACTTCCTGCTGCGCGCCACAGTGC 424
Db 7794688 GCTGCTGGAATCCCCCTCGCACAGTACGAGTGGCCGCTCCCTGCTGCGCTGCTGCG 7794747

Qy 425 CATCCGCGTG----CGCGAGGGCTCAGAGGCCACCTTCGCTGCGCGCTGGGTGGCTCCC 480
Db 7794748 CGCCCTCGCGAGTACGCGGAGGCTCCCGCGAGCTTCAGCTGGCGCGCGCGGATGCT 7794807

Qy 481 CGAGGCGCGAGTGGTGGTCCAGGAGCGGGCGCGCTGGGTGAGCCGACGCGGCCCGCC 540
Db 7794808 CGA-GCGGACCGTGGCGGTGCTGCACTTCGCGGGCCATCACCGAGACGAGCATCG 7794866

Qy 541 GCGTGGCGGTGAGGAGCTCGCGAGGCAAGTTCGCTGGCGCATTCGCGGGCGCGCGCGCG 600
Db 7794867 AGCGCGCGGAGAGCTCTCGGGATGCCCCGCGACGCTGCGCTTCGCCGAAACCCCGCGCG 7794926

Qy 601 GCGACGCGCGCACTTACAGGCTCGCGCGCGAGAACCGGCTGGCGGCTGCGAGCGCGCGCG 660
```


FEATURE:
NAME/KEY: misc_feature
LOCATION: (390)..(390)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (392)..(393)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (396)..(417)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (419)..(420)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (422)..(426)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (429)..(430)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (433)..(433)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (435)..(435)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (437)..(438)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (441)..(441)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (447)..(448)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (450)..(450)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (453)..(455)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (460)..(460)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (463)..(466)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (470)..(471)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (474)..(474)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (478)..(480)

Query Match 7.0%; Score 74.6; DB 13; Length 1026;

Best Local Similarity 31.8%; Pred. No. 4.7e-08;
Matches 200; Conservative 0; Mismatches 429; Indels 0; Gaps 0;
QY 184 CGCCACAGGTGAGCTGGGAGAAAGGACAGACCGGTGACGCCGCGCGCGCTTCGGTC 243
Db CACCNCGCGGGGGGNGGNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 431
QY 244 TGGCCACAGACGGCGACCTCTACCGCTCACTATCTTGACCTGGCGCTGGCGACAGTG 303
Db GNGCNCCNCCCGGNNCCNNNGGCGNCCNNNNCGGNGGNGGNGGNGGNGGNGGNGGNGG 491
QY 304 GGCATATACGTGTGCGCGCGCAATGCCATAGGCGAGGCTTCGCTGCGTGGGCTTCG 363
Db NCNNGNNGGCGNNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 551
QY 364 AGTGGACCGGAGGCGCGTGGCGGAGCAGCGCGCGCTTCCTGCTGCGGCGCACTG 423
Db NGNNGGNGGCGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 611
QY 424 CCATCGCGTGGCGAGGCTCAGAGGCGCACCTTCGCTGCGCGCTGGGTGGCTCCCGA 483
Db CGCGGCGCGGNTCCNCCGCGGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 671
QY 484 GCGCGGACGTGAGCTGCTCAAGGACGGCGCGGCTGGGTGAGCCGCGCGCGCGCGG 543
Db NGGNGGNGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 731
QY 544 TGGCGGTGAGGAGCTCGGCGAGGCAAGTGCGCTGCGCATTCGGCGCGCGCGCGCG 603
Db GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 791
QY 604 ACGGCGGCACTTACGAGGTCCGCGCGAGAAACCGCTGGCGCTGCGAGCGCGCGCGG 663
Db CCNGCGAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 851
QY 664 CGCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 723
Db CCNNGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 911
QY 724 CGCTCTGCGCGACCTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 783
Db GCGCGNTNCACTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 971
QY 784 CACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 812
Db NNTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1000

Search completed: September 20, 2004, 02:51:45
Job time : 572.829 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:16:23 ; Search time 92 Seconds
(without alignments)
4471.258 Million cell updates/sec

Title: US-10-077-130-5
Perfect score: 41273
Sequence: 1 MDQFSGARFLTRPKAFV.....RNREKRALLYKEHNLQVR 7968

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pdp.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pdp.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pdp.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pdp.*
5: /cgn2_6/ptodata/2/iaa/ECTUS.COMB.pdp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 8423 | 20.4 | 1665 | 4 | US-09-858-664A-2 |
| 2 | 8423 | 20.4 | 1665 | 4 | US-10-274-978-2 |
| 3 | 1173 | 2.8 | 846 | 4 | US-09-858-664A-3 |
| 4 | 1173 | 2.8 | 846 | 4 | US-10-274-978-4 |
| 5 | 909.5 | 2.2 | 549 | 4 | US-09-858-664A-5 |
| 6 | 909.5 | 2.2 | 549 | 4 | US-10-274-978-6 |
| 7 | 892.5 | 2.2 | 2860 | 2 | US-08-826-267-2 |
| 8 | 592.5 | 1.4 | 414 | 4 | US-09-858-664A-13 |
| 9 | 592.5 | 1.4 | 414 | 4 | US-10-274-978-14 |
| 10 | 531 | 1.3 | 279 | 4 | US-09-858-664A-4 |
| 11 | 531 | 1.3 | 279 | 4 | US-10-274-978-5 |
| 12 | 521 | 1.3 | 298 | 4 | US-09-858-664A-17 |
| 13 | 521 | 1.3 | 298 | 4 | US-10-274-978-18 |
| 14 | 511.5 | 1.2 | 508 | 4 | US-09-858-664A-18 |
| 15 | 511.5 | 1.2 | 508 | 4 | US-10-274-978-19 |
| 16 | 504 | 1.2 | 250 | 4 | US-09-858-664A-6 |
| 17 | 504 | 1.2 | 250 | 4 | US-10-274-978-7 |
| 18 | 489 | 1.2 | 448 | 3 | US-09-159-385-2 |
| 19 | 489 | 1.2 | 448 | 3 | US-09-186-277-2 |
| 20 | 476.5 | 1.1 | 11877 | 3 | US-09-105-537-6 |
| 21 | 472.5 | 1.1 | 260 | 2 | US-07-857-224B-23 |
| 22 | 466.5 | 1.1 | 454 | 2 | US-09-159-385-1 |
| 23 | 466.5 | 1.1 | 454 | 2 | US-09-186-277-1 |
| 24 | 466 | 1.1 | 274 | 4 | US-09-858-664A-14 |
| 25 | 466 | 1.1 | 274 | 4 | US-10-274-978-15 |
| 26 | 465 | 1.1 | 358 | 4 | US-09-230-896C-29 |
| 27 | 464.5 | 1.1 | 1050 | 4 | US-09-428-711A-16 |

28 456 1.1 331 3 US-08-810-712-24 Sequence 24, Appl
29 456 1.1 1423 3 US-08-810-712-10 Sequence 10, Appl
30 454.5 1.1 356 4 US-09-733-388-4 Sequence 4, Appl
31 454 1.1 1651 3 US-09-540-245A-18 Sequence 18, Appl
32 451.5 1.1 1051 4 US-09-428-711A-14 Sequence 14, Appl
33 446 1.1 424 2 US-08-715-568A-1 Sequence 1, Appl
34 438.5 1.1 385 4 US-09-733-388-2 Sequence 2, Appl
35 436 1.1 355 4 US-09-579-664B-10 Sequence 10, Appl
36 433 1.0 261 2 US-07-857-224B-22 Sequence 22, Appl
37 432 1.0 343 3 US-08-878-989-5 Sequence 5, Appl
38 432 1.0 343 3 US-09-272-796-5 Sequence 5, Appl
39 427 1.0 370 2 US-08-878-989-19 Sequence 19, Appl
40 427 1.0 370 3 US-09-272-796-19 Sequence 19, Appl
41 427 1.0 370 4 US-09-457-040B-31 Sequence 31, Appl
42 418.5 1.0 1395 3 US-09-540-245A-15 Sequence 15, Appl
43 413.5 1.0 307 1 US-08-713-828-1 Sequence 1, Appl
44 413.5 1.0 307 2 US-08-918-627-1 Sequence 1, Appl
45 413.5 1.0 307 2 US-09-096-245-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-858-664A-2
; Sequence 2, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-2

Query Match 20.4%; Score 8423; DB 4; Length 1665;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6356 PPSMQVTIEDVQAOTCGTAQFAEALLEGDPQPSVTWKDSVOLVDSTRLSQQCGEGTYSLV 6415
DB 53 PPSMQVTIEDVQAOTCGTAQFAEALLEGDPQPSVTWKDSVOLVDSTRLSQQCGEGTYSLV 112
QY 6416 LRHVASKDAGVYTCLAQNTGGQVLCKAELLVLGGDNEPDSEKSHRRKLSHFEVKEEIG 6475
DB 113 LRHVASKDAGVYTCLAQNTGGQVLCKAELLVLGGDNEPDSEKSHRRKLSHFEVKEEIG 172
QY 6476 RGVFGFKVRQVHKGKILCAAKFIPLSRSTRAQVREIRILAAISHPLVTGLDQPTETRK 6535
DB 173 RGVFGFKVRQVHKGKILCAAKFIPLSRSTRAQVREIRILAAISHPLVTGLDQPTETRK 232
QY 6536 TLTLLELCSBELLDRLVRKGVWVBAEVKVIQQLVEGLVHSHGVHLHDIKPSNIML 6595
DB 233 TLTLLELCSBELLDRLVRKGVWVBAEVKVIQQLVEGLVHSHGVHLHDIKPSNIML 292
QY 6596 VHPAREDIKIDGFGAQTTPAELQFSQYGSPEFVSPEIIQQNPVSEASDIWAMGVISYL 6655
DB 293 VHPAREDIKIDGFGAQTTPAELQFSQYGSPEFVSPEIIQQNPVSEASDIWAMGVISYL 352
QY 6656 SLTCSPPFAGESDRATLLNVLEGRVSSPMAHLSSEDAKDFIKATLQAPQAPPSAAQC 6715
DB 353 SLTCSPPFAGESDRATLLNVLEGRVSSPMAHLSSEDAKDFIKATLQAPQAPPSAAQC 412

Db 533 AEAASERTEAPAPSPAGAGPAAQGCVPVHSVIRSLFYHQAGSEPHGALAPGSRHP 592
QY 6896 ARRHLLKGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSEFETALRLPASGTH 6955
Db 593 ARRHLLKGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSEFETALRLPASGTH 652
QY 6956 LAPGSHSLHDSSTPRPSSAEGEAQRLPSAGGAPTRDMGHPOGSKOLSTGCHPG 7015
Db 653 LAPGSHSLHDSSTPRPSSAEGEAQRLPSAGGAPTRDMGHPOGSKOLSTGCHPG 712
QY 7016 TAQPERPSDFPMQOPAFCHPKQGSAPQEGCSPEHPAVAPCTPGSPFPFGCKEAPLVPS 7075
Db 713 TAQPERPSDFPMQOPAFCHPKQGSAPQEGCSPEHPAVAPCTPGSPFPFGCKEAPLVPS 772
QY 7076 PFLQOPAPPAPAKASPLDCKMGPGDISLPGRKPGPCSPGASQASQSSVSLRVS 7135
Db 773 PFLQOPAPPAPAKASPLDCKMGPGDISLPGRKPGPCSPGASQASQSSVSLRVS 832
QY 7136 SOVGTEPGPSDAEGWTOEADLSDSTPTLORPOQOATMRKFSLGGRGGVAGVAGYTGA 7195
Db 833 SOVGTEPGPSDAEGWTOEADLSDSTPTLORPOQOATMRKFSLGGRGGVAGVAGYTGA 892
QY 7196 FGGDAGMLGGQPMWARIAWAVSQSEEEQBEAASQSEEQBARAESPLPQVSAAPVP 7255
Db 893 FGGDAGMLGGQPMWARIAWAVSQSEEEQBEAASQSEEQBARAESPLPQVSAAPVP 952
QY 7256 EVGRAPTRSSPEPTWEDIGOVSLVOIRDLSCDAEADTISLDISEVDPRAYLNLSLDYDI 7315
Db 953 EVGRAPTRSSPEPTWEDIGOVSLVOIRDLSCDAEADTISLDISEVDPRAYLNLSLDYDI 1012
QY 7316 KYLPPEFMI FRKVPKSAQPEPPSWAEBELAEFFPEPTWMPWPGELGPHAGLEITESESDVD 7375
Db 1013 KYLPPEFMI FRKVPKSAQPEPPSWAEBELAEFFPEPTWMPWPGELGPHAGLEITESESDVD 1072
QY 7376 ALLAEAAVGRKWKSSPSRSLFHPFGRHLPLDEPAELGLRRVKAASVEHISRLIKGRPEG 7435
Db 1073 ALLAEAAVGRKWKSSPSRSLFHPFGRHLPLDEPAELGLRRVKAASVEHISRLIKGRPEG 1132
QY 7436 LEKEGPRKKGGLASFRLSGLKSWDRAPTFLRELSDETIVLGQSVTLACQVSAQPAQAAT 7495
Db 1133 LEKEGPRKKGGLASFRLSGLKSWDRAPTFLRELSDETIVLGQSVTLACQVSAQPAQAAT 1192
QY 7496 WSKGAPLESSRVLISATLNKFNOLLITLVVAEDLGVTTCVSNALGTVTTCGLRKA 7555
Db 1193 WSKGAPLESSRVLISATLNKFNOLLITLVVAEDLGVTTCVSNALGTVTTCGLRKA 1252
QY 7556 RPSSSPCFDICEVADGVLLVWKPVESYGPVTYIVQCSLEGSWTTLASDIFDCCYLTSK 7615
Db 1253 RPSSSPCFDICEVADGVLLVWKPVESYGPVTYIVQCSLEGSWTTLASDIFDCCYLTSK 1312
QY 7616 LSRGTYTFRACVSKAGMPYSSPSEQVLLGGPSHLASEESQGRSAQPLPSTKTFAPQ 7675
Db 1313 LSRGTYTFRACVSKAGMPYSSPSEQVLLGGPSHLASEESQGRSAQPLPSTKTFAPQ 1372
QY 7676 TQIQGRSVVROCKEASGALAAKIIPHPKDKTAVLREYEAALKGLRHPHLAQLHAAY 7735
Db 1373 TQIQGRSVVROCKEASGALAAKIIPHPKDKTAVLREYEAALKGLRHPHLAQLHAAY 1432
QY 7736 LSPRLVLI LELCSGPELLPCLAEASYSSEVKDYLWQMLSATQYLHNQHILHLDRSE 7795
Db 1433 LSPRLVLI LELCSGPELLPCLAEASYSSEVKDYLWQMLSATQYLHNQHILHLDRSE 1492
QY 7796 NMITEYNLLKVDIGNAQSLSQKVLPSDFKDYLETMAPELLEGQGVQPTDIWAGV 7855
Db 1493 NMITEYNLLKVDIGNAQSLSQKVLPSDFKDYLETMAPELLEGQGVQPTDIWAGV 1552
QY 7856 TAFIMLSAEPVSSSEGADLQGLRKGLVRLSRVCAGLSGGAFLASTCAQPWGRPCA 7915
Db 1553 TAFIMLSAEPVSSSEGADLQGLRKGLVRLSRVCAGLSGGAFLASTCAQPWGRPCA 1612
QY 7916 SSCLOCPWLTBEGPACSRPAPVTFTARLRVVRNREKRALLYKHNLAQVR 7968

Db 1613 SSCLOCPWLTBEGPACSRPAPVTFTARLRVVRNREKRALLYKHNLAQVR 1665

RESULT 3
US-09-858-664A-3
; Sequence 3, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 846
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-858-664A-3

Query Match 2.8%; Score 1173; DB 4; Length 846;
Best Local Similarity 34.9%; Pred. No. 1,1e-54;
Matches 313; Conservative 121; Mismatches 316; Indels 146; Gaps 29;

QY 6357 PSMQVTVI DVQATGCTA QFAEAIIEGDPQPSVTWKDSVQLVDSTRLSQQQGGTYSLVL 6416
Db 1 PRFESIMEDVEVGAGTARFAVVGKPLDPMWYKDEVLITSSHSVSVSENECSLVV 60
QY 6417 RHVASDKAGVYTCIAQNTGGVLCCKAELLVLGGDN---EPDEKQSHR-RKLHSYEVK 6471
Db 61 LSTGAQGGVYCTAQNLAGVSCAKELAVHQAQTAMEVEGVEDEHGRRLSDFYDIH 120
QY 6472 EIEGRGVGFVKRVQKGNKILCAAKFIPLRSRTAQAYRERDILAAALSHPLVTGLDQF 6531
Db 121 QEIGRGAFSYLRIRIVERSSGLEFAAKFIPSQAKPKASAREARLLIARLQHDCLVYFHEAF 180
QY 6532 ETRKTLILILECSSBELDLRYKGVVTEAEVKYIQQVLEGLHYLHSHGVHLHDIKPS 6591
Db 181 ERRRGVLIVTELC-TEELLERJARKPTVCESEIRAYMRQVLEGIHYLHSHGVHLHDVKE 239
QY 6592 NILMVHFA--REDIKICDFGFAQNTITPAELQFSQSGSPFVSPETIQNPVSEASDIWAM 6649
Db 240 NLLVMDGAAGEQVVRICDFGNAQELTPGEPQICQGTPEFAVEIYNQSPVSGVTDMFV 299
QY 6650 GVISYLSLTCSPFPAGESDRATLLNVLEGRVSWSPMAHLSEDAKDF-IKATLQAPQA 6708
Db 300 GVVAFLCLTGISPFVENDRTTLMNIRYNVAFEETTFLSLSREARGFLIKVLVQ--DRL 357
QY 6709 RPSAAQCLSHPWFLKSMPEAEAHFINTKOLKFLARSWRORSIMSYKSLVMRSIPELLR 6768
Db 358 RPTAETLEHPFKTQAKGAE--VSTDHLKLFISRRRWORSQISYKCHLVLRPPELLR 414
QY 6769 GPDPSPSLGVARHLCDTTCGSSSSSSSSDNEL-----APPARAK-SLPSPSVTH 6816
Db 415 APPERVVMTMPRR-PPPSGGLSSSDSEEELEELPSVPRPLQPEFSGSRVSLDITPDT 473
QY 6817 SPLHPRGFLRPSASLPEEAASEASERSTEAPAPASPEGAGPAAAGCVPVHVR----- 6871
Db 474 EALGTPETGAATMDMQEQGRAPSQOQEAPEALLPSPQGEFAA-CASPRRELARGSSA 532
QY 6872 -----SLFYHQAG--ESPEHGALAPG-----SRHPARARHLK 6903
Db 533 ESALPRAGPRELGRGLHKAASVELPQRRSPGFGATLARGGLGEGEYAQLQALQRLR 592
QY 6904 GGVIAGALPGLREPLMEH-----RVLEEAAREEQATL-----LAKAPSETALR 6948
Db 593 GCPEDKVGSLRGPLLESLGGRAPDRVARAAASAEAPHHPQPLENRLGLOKSSSQSGBA 652

QY 6949 LPASGTHLAPGHSHLSHDSPTPR-----PSSACGGAQRLPSAPSGGAPIRDMGHPOGS 7004
 Db 653 EP-RGRHRRAGAPLEIPVARLGARLQESPSLSALSEAQ--PSSPA-----RPSAP 700
 QY 7005 KQLPSTGCHPGCTAOPERPSPDWPQAPFCHPKQGSAPQEGCSPHPAVAPCPGSPFPFG 7064
 Db 701 K--PST---PKSABESATTSDAPQPPAP--QPAQDKAPEPRPVPVASKPAP-----PPQ 749
 QY 7065 SCK--EAPLVPSPPFLOQ-----PQAPPAPAK-----ASPPLDSKMG 7100
 Db 750 ALQTLALPLTPTVAQIIQSLQSLGSHAQGPSQCPAAPPSEPKPFAAFAVAVASPP-----802
 QY 7101 GDISLGRPKPGPCSSPGSASQASSQVSSLRVSSOVGTGPGPSLDAEGWTQEA 7156
 Db 803 -----PGAPEKRVPSAGGPPVLAEKARVPT-----VPPRPGSSLSSENLESE 846

RESULT 4
 US-10-274-978-4
 ; Sequence 4, Application US/10274978
 ; Patent No. 6670164
 ; GENERAL INFORMATION:
 ; APPLICANT: WEI, Ming-Hui, et al
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
 ; TITLE OF INVENTION: THEREOF
 ; FILE REFERENCE: CL000927-CIP-DIV
 ; CURRENT APPLICATION NUMBER: US/10/274,978
 ; CURRENT FILING DATE: 2002-10-22
 ; PRIOR APPLICATION NUMBER: 09/858,664
 ; PRIOR FILING DATE: 2001-05-17
 ; PRIOR APPLICATION NUMBER: 09/711,134
 ; PRIOR FILING DATE: 2000-11-14
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 846
 ; TYPE: PRT
 ; ORGANISM: Human
 US-10-274-978-4

Query Match 2.8%; Score 1173; DB 4; Length 846;
 Best Local Similarity 34.9%; Pred. No. 1.1e-54;
 Matches 313; Conservative 121; Mismatches 316; Indels 146; Gaps 29;

QY 6357 PSMQVTT-EDVQAQGGTAQFEAIEGDPQPSVTWKDVSQVLDSTRLSQOQEGTYSVLV 6416
 Db 1 PRFSIMEDVEVGAGETARFAVVVEGKPLDIMWKDEVLLTSSHSVFVYENECSLVV 60
 QY 6417 RHVASKDAGVYTCLAQNTGGVLCABELLVLGGDN-----EPDSEKOSHR-RKLHSFYEVK 6471
 Db 61 LSTGAQDGGVYTCTAQNLAGESVCKAELAVHSAQTAMEVEGVGEDEHGRRLSLDFYDIH 120
 QY 6472 BEICRGVGFVKRYQVHKGNTLCAAKFIPLRSTRQAQYRERDILAAISHPLVTGLLDOF 6531
 Db 121 QEIGRGAFYLRIVERSSGLEFAKFIPOAKPKASAREARLLARLQDCHVLYPHEAF 180
 QY 6532 ETRKTLILIELCSBELDLRYKGVVTEAEVKYIQQLVEGHYHLHSHVHLIDKPS 6591
 Db 181 ERRRGLVIVTELC-TEELLERIAKPTVCESEIRAYNRVLEGITHYLHSHVHLIDVQKE 239
 QY 6592 NILMVHRA--REDIKICDFQAQNTIPAELOFQSGSPFVPSPEILQONPVSEASDIWAM 6649
 Db 240 NLLVWDGAGGQQVKRICDFGNAQELTFCPEQYCOYGTPEFAVEINQSPVSGVTIDWPV 299
 QY 6650 GVISYLSLTCSPPAGESDRATLLNVLEGRVSSPMAAHLSEDAKDF-IKATLQAPQA 6708
 Db 300 GWAFCLCTGISPPFVGENDRITLLNRYNVAPEETFLSLREARGFLIKVLVQ--DRL 357
 QY 6709 RPSAAQCLSHWFLKSPAEAEHINTKQLKFLARSRWORSLSYKSIIVVRSIDPELLR 6768
 Db 358 RPTAEETLEHFWFKTQAKGAE--VSTDHLKFLSRRRWORSQISYKCHLVLRPIPELLR 414

QY 6769 GPPDSFSLGVARHLCDRTDGGSSSSSSSSSDNEL-----APPAAK-SLPPSPVTH 6816
 Db 415 APPERVVWVTPRR-PPPSGGLSSSSDEEELEBLPSVPRFLQPEFGSRVSLTDIPTED 473
 QY 6817 SPLLPGRGFLRPSASLPEAEASERSSTEAPAPPASPEGAGPPAAQGVCPRHSHVIR-----6871
 Db 474 EALGTPTGTAATPMDWQEGCRAPSDQDQAPSPALPSQGEPA--GASPRGELRGSSA 532
 QY 6872 -----SLFYHQAG--ESPEHCALAPG-----SRHPARRHLK 6903
 Db 533 ESALPRAGPRELGRGLHKAASVLPQRRSPGPGATRLARGGLGEGEYAOQLAQLRQLLR 592
 QY 6904 GGIYIAGALPGRLPELMEH-----RVLEEAAREEQATL-----LAKAPFFETALR 6948
 Db 593 GGPEDGVKVGRLGRLLESGLGRARDPVARAASSEAAHPHQPLENGLQKSSFSQGEA 652
 QY 6949 LPASGTHLAPGHSHLSHDSPTPR-----PSSACGGAQRLPSAPSGGAPIRDMGHPOGS 7004
 Db 653 EP-RGRHRRAGAPLEIPVARLGARLQESPSLSALSEAQ--PSSPA-----RPSAP 700
 QY 7005 KQLPSTGCHPGCTAOPERPSPDWPQAPFCHPKQGSAPQEGCSPHPAVAPCPGSPFPFG 7064
 Db 701 K--PST---PKSABESATTSDAPQPPAP--QPAQDKAPEPRPVPVASKPAP-----PPQ 749
 QY 7065 SCK--EAPLVPSPPFLOQ-----PQAPPAPAK-----ASPPLDSKMG 7100
 Db 750 ALQTLALPLTPTVAQIIQSLQSLGSHAQGPSQCPAAPPSEPKPFAAFAVAVASPP-----802
 QY 7101 GDISLGRPKPGPCSSPGSASQASSQVSSLRVSSOVGTGPGPSLDAEGWTQEA 7156
 Db 803 -----PGAPEKRVPSAGGPPVLAEKARVPT-----VPPRPGSSLSSENLESE 846

RESULT 5

US-09-858-664A-5
 ; Sequence 5, Application US/09858664A
 ; Patent No. 6482624
 ; GENERAL INFORMATION:
 ; APPLICANT: WEI, Ming-Hui, et al.
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
 ; TITLE OF INVENTION: THEREOF
 ; FILE REFERENCE: CL000927-CIP
 ; CURRENT APPLICATION NUMBER: US/09/858,664A
 ; CURRENT FILING DATE: 2001-05-17
 ; PRIOR APPLICATION NUMBER: 09/711,134
 ; PRIOR FILING DATE: 2000-11-11
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 549
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-858-664A-5

Query Match 2.2%; Score 909.5; DB 4; Length 549;
 Best Local Similarity 38.3%; Pred. No. 8.1e-41;
 Matches 204; Conservative 88; Mismatches 181; Indels 59; Gaps 12;

QY 6274 TGTSEAPVPRVPPQPLLHEGPECEPAIARAQEWTVPIRMEGAAMPFAGT-----GE 6326
 Db 25 TVKSSSKSPSEFPVQLLEHGFTLE-EAPAMLDKPDIVVVEGQ--PASVTVTENHVEAQ 81
 QY 6327 LLVDVHSHVRETQRTTYQAID-----THTAR-----6355
 Db 82 VVWRSCRGALLAEARAGVYELSQDDQYCLICRVSRDRCALTCIARNHGTCTCSTVL 141
 QY 6356 ----PPSMQVITIEDVQAQGTGTAQFEAIEGDPQPSVTWKDVSQVLDSTRLSQOQEGT 6411
 Db 142 ELAEAPRESIMEDVEVGAGETARFAVVVEGKPLDIMWKDEVLLTSSHSVFVYENE 201
 QY 6412 YSLVLRHVASKDAGVYTCLAQNTGGVLCABELLVLGGDN-----EPDSEKOSHR-RKLHS 6466

Db 202 CSLVVLSTGADGGVYTCTAQNLAGVSCXAEALVHSAQTAAMEVGVGEDEHRRRLSD 261
Qy 6467 FVEYKEBIGRGVGFVQRVQKGNKILCAAKFIPLRSRTAQAYRREDILAAALSHPLVTG 6526
Db 262 FYDIHQBIGRGAFSYLRIVERSSGLEFAAKFIPSOAKPKASAREARLLARLQHDVCVLY 321
Qy 6527 LLDQFETRKTLILILELCSSEELLDRLYRKGVVTEABVKYIQOVLGELHSHGVHLH 6586
Db 322 FHEAFERRRGLVIVTELC--TEELLERIAKPTVCESEIRAYMRQVLEGIHVLHSHVHLH 380
Qy 6587 DIKPSNIMLVHFA--REDIKICDFGFAQNTPAELQFSQYSGSPFVSPFIIQONPVSEAS 6644
Db 381 DVKENLLVWDGAAGEQQVRIICDFGNAQELTPGEPQYCYGTPEFVAPEIVNQSPVSGVT 440
Qy 6645 DIWANGVLSYLSLTCSSPPAGESDRATLLNVLEGRVSWSSPMAHLSEDADKF-1KATLQ 6703
Db 441 DIWPGVGVAFCLTGISPFVGENDRITLMNRNYNVAFEETFTLSLSREARGFLIKVLVQ 500
Qy 6704 RAPQARPSAAQCLSHPWFLKSMPEAEAHFINTKOLKELLARSWRQSRMSYK 6755
Db 501 --DLRPTAETLEHPWFKTQAKGAE--VSTDHLKFLSRRRWQSRQISYK 547
RESULT 6
US-10-274-978-6
; Sequence 6, Application US/10274978
; Patent No. 6670164
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CLC00927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Human
US-10-274-978-6
Query Match 2.2%; Score 909.5; DB 4; Length 549;
Best Local Similarity 38.3%; Pred. No. 8.1e-41;
Matches 204; Conservative 89; Mismatches 181; Indels 59; Gaps 12;
Qy 6274 TGTSEAPVPRVPQPLIHGEGEQEPEALARAQEWTPIRMEGAAWFGAGT-----GE 6326
Db 25 TVKSSSKPSPSEPQVLLHGHGTTLE-EAPALMDKPDIVYVVEGQ--PASVTVTFNHVEAQ 81
Qy 6327 LLMDVHSHVRETQRTTYQAID-----THTAR----- 6355
Db 82 VVMRSRGALLAEARAGVYELSQPDDQVCLRICVSRDRMGALCTARNRHGTQTCVTL 141
Qy 6356 ----PPSMQVITEDVQACTGTAQFAEIIEDGPOPSVTWYKDSVOLVDSTRLSQQQSGTT 6411
Db 142 ELAEPAPRESIMEDVEVCAGETAFVAVVEGKPLPDINWYKDEVLITESHSHVSFVLENE 201
Qy 6412 YSLVLRHVASKDAGVYTCLANTGGQVLCCKAELLVLGGCN----EPQSEKQSHR-RKLHS 6466
Db 202 CSLVVLSTGADGGVYTCTAQNLAGVSCXAEALVHSAQTAAMEVGVGEDEHRRRLSD 261
Qy 6467 FVEYKEEIRGRGVGVKGVQKGNKILCAAKFIPLRSRTAQAYRREDILAAALSHPLVTG 6526
Db 262 FYDIHQBIGRGAFSYLRIVERSSGLEFAAKFIPSOAKPKASAREARLLARLQHDVCVLY 321
Qy 6527 LLDQFETRKTLILILELCSSEELLDRLYRKGVVTEABVKYIQOVLGELHSHGVHLH 6586

Db 322 FHEAFERRRGLVIVTELC--TEELLERIAKPTVCESEIRAYMRQVLEGIHVLHSHVHLH 380
Qy 6587 DIKPSNIMLVHFA--REDIKICDFGFAQNTPAELQFSQYSGSPFVSPFIIQONPVSEAS 6644
Db 381 DVKENLLVWDGAAGEQQVRIICDFGNAQELTPGEPQYCYGTPEFVAPEIVNQSPVSGVT 440
Qy 6645 DIWANGVLSYLSLTCSSPPAGESDRATLLNVLEGRVSWSSPMAHLSEDADKF-1KATLQ 6703
Db 441 DIWPGVGVAFCLTGISPFVGENDRITLMNRNYNVAFEETFTLSLSREARGFLIKVLVQ 500
Qy 6704 RAPQARPSAAQCLSHPWFLKSMPEAEAHFINTKOLKELLARSWRQSRMSYK 6755
Db 501 --DLRPTAETLEHPWFKTQAKGAE--VSTDHLKFLSRRRWQSRQISYK 547
RESULT 7
US-08-826-267-2
; Sequence 2, Application US/08826267
; Patent No. 5994070
; GENERAL INFORMATION:
; APPLICANT: Streuli, Michel
; TITLE OF INVENTION: No. 5994070e1 TRIO Molecules and Uses Related Thereto
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: LAHIVE & COCKFIELD
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,267
; FILING DATE: 1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,214
; FILING DATE: 27 MARCH (1996)
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2860 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-826-267-2
Query Match 2.2%; Score 892.5; DB 2; Length 2860;
Best Local Similarity 24.7%; Pred. No. 6.9e-39;
Matches 318; Conservative 209; Mismatches 493; Indels 265; Gaps 42;
Qy 5562 DSDSDSKTSPASPRHGRSR----PSSIGSESSSESD--GDARGEIFDIYVVTADYLP- 5614
Db 1758 DSDSDSAATQDFTVEERGNGELSGTILSKSSSGMQSCGEEG-----EGADAVPLP 1811
Qy 5615 ---GAPQDAITREGQYVEVLDAHPRLWLVTKTKTKS--SPSRQGWSP-----AY 5661
Db 1812 PPMAIQOHSLLQPDQSQ----DDKASSRLVR--PTSETPSAABELVSAIBELVSKMAL 1864
Qy 5662 LDRLRLKLSPEWGAAPAF-PGE-----AVSEDEYKAEKLS-----VIGELLSSEGA 5707
Db 1865 EDRPSSLVDQGDSSSPFNPSDNLSSSPIDEMERKSSSLKRRHVVLOELVETED 1924

| | | |
|----|------|---|
| Qy | 5708 | FVELOFLOSHHLOHLERCHPV:AVAGOKAVI:FNVRDITGRFHS-:FLOELQOCDTD-D:576 |
| Db | 1925 | YRDLGVVVEGYMA-LMKEDGVPDMRGKOKIVIFGNHIOHYDMHRDFFLGELCKLEDPE:1983 |
| Qy | 5766 | DVAMCFIKQAAFAEOYLEFLVGRVOAESVVVSTAIQBFYKYAEEALLAGDPSPPPPPPL:5825 |
| Db | 1984 | KLGSIFVKHERLHMYTAYCQNKPKSEH-IVSEVIDTFPDLKOR-----L:2028 |
| Qy | 5826 | QHYLE-----QPVERVOROALLKEILIRNKARNRQNCALLEQAYAVVSALPORAENKCH:5879 |
| Db | 2029 | CHRLQLDOLLKPVORIMKYQALLKDFLKYSKASLDTSELAVERAVMCI:VPRCNDMMN:2088 |
| Qy | 5880 | VSLMENPGTLEALGEPITROGHEFTVWEGCAPRWKGNHRHVFLEFNHLVICRPRDSR:5939 |
| Db | 2089 | VGRLOGGDGKIVAOGKULLQDTFLVTDQDAGL-LP-RCRERRIFLFOIVIFSEPLDKKX:2146 |
| Qy | 5940 | TDTV-SYVFRNMKLSIDLNDQVEGDDRAFEVWQEREDSVRKYLLOARATAIKSSWYKE:5998 |
| Db | 2147 | GFSMPGELFKNSIKVSCLCLEENVENDPCKFALTSRTGDVVETFILHSSSPSVRQTWIHE:2206 |
| Qy | 5999 | ICGI---QORLALFVWRPPPEEBL-----ADCTAELGETVKLACRTVGTGPKVISW:6047 |
| Db | 2207 | INQILEQORNFLNALTSPIEYQRNHSGGGGGSGAAGVG-----AAAAAGPFAAAAT:2261 |
| Qy | 6048 | YKDGKAVQVDPHILIEDPDGSCALILDSLTGVDSGQYMCFAASAAGNCSTGLKILVQVP:6107 |
| Db | 2262 | VAAAPAAAAAPP-----ARAGAG-----P:2279 |
| Qy | 6108 | PRFNVKVRASPFVEGEDAQFTCTIEGAPYQPIRWYKDGLTTGNKPTLSEPRSGLLVL:6167 |
| Db | 2280 | P-----GSPSL-SDTTPPCSPLOPRARQR-----QTRCQSSSSSSSISTM:2320 |
| Qy | 6168 | VI---RAASKEDLGLYCEELVNRFLGASASAEILQSPMLQAEQC-----6210 |
| Db | 2321 | LVTHDYTAVEDEINTVQGEVVQILASQNQMLVFPRA-----ATDQC:PAAGHWIPGFVLG:2376 |
| Qy | 6211 | HREQLVAAVEDTTLIERADQEVTSVLKELLGPK--APGFSTGDTLGP:CPRGAPALQETG:6268 |
| Db | 2377 | HTSAVIVENDPDGLKKSTSWHTALRLKSEKDKDGKREGKL-----ENG:2422 |
| Qy | 6269 | SQPVVTGTSAPAVPRVPQPLLHEGFEQBEPAIARAQEWTVPIRMEGAWPAGATGELL:6328 |
| Db | 2423 | YKRSREGLSN---KVSVKLLNP-----2441 |
| Qy | 6329 | WDVHSHVVRETTORTYTYQAIIDHTARTPPSNQVMTIEDVQAQTGTAQFAEAI:EGDPOPSV:6388 |
| Db | 2442 | -----NYIYDV-----PPEFVPLSEVTCETGETVVLRCFVCGCRPKASI:2480 |
| Qy | 6389 | TWYKDSVLVD-----STRLSQOQEGTYSILVRHVASKDAGVYTCLAQNTGQGVLCXAE:6443 |
| Db | 2481 | TWKGPENTLNNDGHYSISYSDLEAT--LKVGVTTEDDGYTCTIAVNDMSGSSASS:2537 |
| Qy | 6444 | LLVLGGDNEPDSKQSHRRKLHGFYEVKEIGRGVFGVKRVQHGKNKILCAAKFTPLRS:6503 |
| Db | 2538 | LRVLGPG--MDGIMVTKDNFDSFYSEVABLGRGRSVVKCKDQGTGRAVATKFNKKL:2595 |
| Qy | 6504 | RTRAQAYREDDILAAALSHPLVTGLLOQFETPKTLILILELCSSEELDLRLYRKGVVTEAB:6563 |
| Db | 2596 | MKRQDVTHELIGLQSLQHPLLVGLLDTFTFTSVILVLEMAQDGRLLDCVWRGSLTEGK:2655 |
| Qy | 6564 | VKVIQOLVEGLVHSHGVLHLDIKPSNLTMLWHP-AREDIKICDGFQAQNTIPALQPS:6622 |
| Db | 2656 | IRAHLGVELAVRYLHNCR:LAHDLKPENILVDESIAKPTIKLADPGDVAQLMNTTYIIHQ:2715 |
| Qy | 6623 | QYGSPEFVSPEIIQONPVSEASDIWAMGVTSYLSLTCSPFPAGESDRATILNVLGERSW:6682 |
| Db | 2716 | LLGNPEFAAPEIILGNPVSUTSDTWSGVLTLYLVSGVSFFLDSVEETCLNICRLDRSF:2775 |
| Qy | 6683 | SSPMAHLSDEADFKATILQAPAPAPPSAAQCISHPWFLKSNPABEAHINTKQLKFLV:6742 |
| Db | 2776 | PDDYFKGVGSKAKFEVCFLOEOPFAKRPASALALQBQW-LQAGNGRSTGVLDTSLRTSTFI:2834 |
| Qy | 6743 | ARSRQW---RSLMSYKSIILVWRGIP:6764 |

```

Db      2835 ERRKHQNDVRPIRSIKNFLOSLRP 2859

RESULT 8
US-09-858-664A-13
; Sequence 13, Application US/09858664A
; Patent NO. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-13

```

```

RESULT 9
US-10-274-978-14
; Sequence 14, Application US/10274978
; Patent No. 6670164
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17

```


;; PRIOR APPLICATION NUMBER: 09/711,134
;; NUMBER OF SEQ ID NOS: 34
;; SOFTWARE: Fast-Seq for Windows Version 4.0
;; SEQ ID NO 14
;; LENGTH: 414
;; TYPE: PRT
;; ORGANISM: Human
US-10-274-978-14

Query Match
Best Local Similarity 1.4%; Score 592.5; DB 4; Length 414;
Matches 145; Conservative 69; Mismatches 191; Indels 15; Gaps 6;

QY 6356 PPSQVITIEDVQACTGTAQFEAIIEDGDPQSVTVYKDSVOLYD----STRLSQQOEGT 6410
DB 1 PPEVIFLSEVTCTGTSTVLRCEVGRPKASITWKQPEHNTLNNDGHYSISYDLSGEAT 60

QY 6411 TYSIVLRHVASKDAGVYTCIAQNTGGQVLCKAELLVLGGNEPDSEKSHRRKLHGFYEV 6470
DB 61 --LKIVGVTTEDDGIYTCIAVNDMGSSASSASLRVLPGG--MDGIMVTWKDNFDSFYSE 115

QY 6471 KEEICRGVGFVXVQHKIKILCAAKFIPLRSETRAQYRERDILAAALSHPLVTGLDQ 6530
DB 116 VAEIGRGVFSYVKKDOKGTRAVATKPVNKKLMKRDQVTHELGILQSLQHPVLVGLD 175

QY 6531 PETRKTLLILELCSBELDLRYKGVVTEAEVYVYIQQLVEGLHYLHSHGVHLIDKP 6590
DB 176 PETFTSVILVLEMAQDQGLLDCVVRWGLSTEGKIRAHLEAVLYLHNCRIAHLDKP 235

QY 6591 SNILMVPH-AREDIKIDCFGAQNTITPAELQFSQYSGPEFVSPIIQQNPVSEASDIWAM 6649
DB 236 ENILVDSLRKPTTKLADFGDAVQLNTYYIHQLGNPEFAAPEIILGNPVSLTSDTWSV 295

QY 6650 GVISVLSITCSSPAGBSDRATLLNVLEGRVSWSSPMAAHLSEDAKDFIKATLQRAQAR 6709
DB 296 GVLTVLLSGVSPFLDDSVETCLNICELDFSPDDYFKGVSKAKSFVCFLLQEDPAK 355

QY 6710 PSAAQCLSHPLKSPMAEPAEAFINTKQLFLARSRWQ---RSLMSYKSLVNRSTPEL 6766
DB 356 PSAAALQEQW-LQAGNGRSTGVLDTSLTSLTSLTSLTSLTSLTSLTSLTSLTSLTSL 414

RESULT 10
US-09-858-664A-4
; Sequence 4, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-4

Query Match
Best Local Similarity 1.3%; Score 531; DB 4; Length 279;
Matches 111; Conservative 45; Mismatches 123; Indels 2; Gaps 1;

QY 7639 SPSEQVLGGPSHLASEESQGRSAQPLPSTTKTFAQTQIQGRFSVVRQCKEASGRAL 7698
DB 1 SPAKEVVSPPGSSPRSSPRPEGTLRQGPQKPYTFLEKARGRFGVVRACRENATGRTF 60

QY 7699 AAKIIPYHPKDKTAVLRVEYALKGLRHPLHLAQLHAAVLSPRHLVILELCSGPPELLPCLA 7758
DB 61 VAKIVPYAAEGKPRVLQVEYEVRLTHHERIMSLHEAVITPRYLVLIAESCGNRELLCGLS 120

QY 7759 ERASYSESEVKDYLMQWLSATQYLHNQHLHLDLRSENMIITEYNLLKVDLGNQAQSLQ 7818
DB 121 DRFRYSEDDVATYVMVQLQGLDYLHGHVHLHDIKPDNLLAPDNALKIVDFGSAQPNP 180

QY 7819 EKVLPDSKFDKDYLETMAPELLEGQGAVPQTDIWAIGTAFIMLSAEYVPSSEGARDLQRG 7878
DB 181 QALRPLGHRGTGLEFMAPEVMKGEPIGSAITDINGAGVLTIMLSGRSPFYEPDPQETEAR 240

QY 7879 LRKGLVRLSRCYAGLSGGAVALFRLSTLCAPQWGRPCASSCL 7919
DB 241 IVGGRFADFQLYPNTSQSATLFLRKVLSVHPWSRP--SSCL 279

RESULT 11
US-10-274-978-5
; Sequence 5, Application US/10274978
; Patent No. 6670164
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Human
US-10-274-978-5

Query Match
Best Local Similarity 1.3%; Score 531; DB 4; Length 279;
Matches 111; Conservative 45; Mismatches 123; Indels 2; Gaps 1;

QY 7639 SPSEQVLGGPSHLASEESQGRSAQPLPSTTKTFAQTQIQGRFSVVRQCKEASGRAL 7698
DB 1 SPAKEVVSPPGSSPRSSPRPEGTLRQGPQKPYTFLEKARGRFGVVRACRENATGRTF 60

QY 7699 AAKIIPYHPKDKTAVLRVEYALKGLRHPLHLAQLHAAVLSPRHLVILELCSGPPELLPCLA 7758
DB 61 VAKIVPYAAEGKPRVLQVEYEVRLTHHERIMSLHEAVITPRYLVLIAESCGNRELLCGLS 120

QY 7759 ERASYSESEVKDYLMQWLSATQYLHNQHLHLDLRSENMIITEYNLLKVDLGNQAQSLQ 7818
DB 121 DRFRYSEDDVATYVMVQLQGLDYLHGHVHLHDIKPDNLLAPDNALKIVDFGSAQPNP 180

QY 7819 EKVLPDSKFDKDYLETMAPELLEGQGAVPQTDIWAIGTAFIMLSAEYVPSSEGARDLQRG 7878
DB 181 QALRPLGHRGTGLEFMAPEVMKGEPIGSAITDINGAGVLTIMLSGRSPFYEPDPQETEAR 240

QY 7879 LRKGLVRLSRCYAGLSGGAVALFRLSTLCAPQWGRPCASSCL 7919
DB 241 IVGGRFADFQLYPNTSQSATLFLRKVLSVHPWSRP--SSCL 279

RESULT 12
US-09-858-664A-17
; Sequence 17, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

QY 6717 SHPWFLK 6723
||| :
Db 501 QHPWLMK 507

RESULT 15

US-10-274-978-19
; Sequence 19, Application US/10274978
; Patent No. 6670164
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: CLC00927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Human
US-10-274-978-19

Query Match 1.2%; Score 511.5; DB 4; Length 508;
Best Local Similarity 25.7%; Pred. No. 1.7e-19;
Matches 125; Conservative 75; Mismatches 170; Indels 117; Gaps 5;
QY 6354 ARPPSMQVTEEDVQAQTGTAQTAQFAIIEGDPQPVWYKQSVQLVSTRLSQOQEGTYS 6413
Db 21 AMPQLIQPEDQKVRAGESVELFGKVTGTPTCTWKFQIQDSEHIKVENSENGSK 80
QY 6414 LVLRHVASKAGVYVTCIAQNTGGQVLCKAEILLVG----- 6448
Db 81 LTLAARQEHGCGCYTLLENKLGSRQAOVNLTVVDKDPDPAGTPCASDIRSSSLTWSYG 140
QY 6449 -----GDNE 6452
Db 141 SSYDGSVAQSVSIEIWDANKWKELATCRSTSFNVQDLLPDHEYKFRVRAINVTSE 200
QY 6453 P-----DSEKQ-----SHRRKLHSFYEVKEEIGRV 6478
Db 201 PSQESLTVGKPEEPKQKWRQCQTDEKPEVDYRTVTINTEQKVSDFYDIERLGS GK 260
QY 6479 FGFVXRVQHGKNIKILCAAKFIPLRSRTRAQVR-ERDILAALSHPLVTGLLDQFETRTL 6537
Db 261 FQOVFLVEKTKRWAGKFFKAYSACEKENIRQEI SIMNCLHHPKLVQCVDAFEERANI 320
QY 6538 ILILELCSSEELDLRYKGV-VTEAEVKVYIQQLVEGLHYLHSHGVHLHDIKFSNIMLV 6596
Db 321 VMVLEIVSGGELFERIIDDFELTERECIKYMRQISEGVYHKQGI VHLDLKPENINCV 380
QY 6597 HPAREDIKICDFGAQNTPAELQSQYSGSPREFVSPELIQQNVSEASDIWANGVVISYLS 6656
Db 381 NKTGTRIKUIDFGLARLENAGSLKVLGTFPEFAPEVINYEPISYATDMWSIGVICYL 440
QY 6657 LTCSSPFPAGESDRATILNLEGRVSWSPMAHLSDEKADFIKATLQAPQAPSAQCL 6716
Db 441 VSLSPFPFGNDNDNETLVNVT SATWDFDDEAFDEISDDAKDFISNLLKKDMKNRLDCTQCL 500
QY 6717 SHPWFLK 6723
||| :
Db 501 QHPWLMK 507

Search completed: September 13, 2004, 11:37:13
Job time : 98 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 19:28:00 ; Search time 199.275 Seconds
(without alignments)
9751.690 Million cell updates/sec

Title: US-10-077-130-4_COPY_16862_17246

Perfect score: 385
Sequence: 1 ccagggcgagatttgacaa.....acctgcagcactggagcgc 385

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|---------------------|
| 1 | 385 | 100.0 | 7893 | 14 | US-10-077-130-2 |
| 2 | 385 | 100.0 | 7928 | 15 | US-10-307-019-5 |
| 3 | 385 | 100.0 | 8106 | 14 | US-10-077-130-1 |
| 4 | 385 | 100.0 | 23907 | 14 | US-10-077-130-6 |
| 5 | 385 | 100.0 | 24120 | 14 | US-10-077-130-4 |
| 6 | 377 | 97.9 | 628 | 11 | US-09-864-408A-1271 |
| 7 | 47.4 | 12.3 | 1290 | 15 | US-10-156-761-193 |
| 8 | 47.4 | 12.3 | 9025608 | 15 | US-10-156-761-1 |
| 9 | 42.6 | 11.1 | 4180 | 13 | US-10-342-887-1266 |
| 10 | 42.6 | 11.1 | 4180 | 13 | US-10-172-118-1266 |
| 11 | 42.6 | 11.1 | 4180 | 16 | US-10-439-388-10 |
| 12 | 42.6 | 11.1 | 4180 | 17 | US-10-776-827-93 |
| 13 | 42.6 | 11.1 | 4282 | 9 | US-09-778-927A-16 |
| 14 | 42.6 | 11.1 | 4366 | 14 | US-10-044-090-668 |

15 41.4 10.8 1803 9 US-09-965-528-46 Sequence 46, Appl
16 41.4 10.8 1803 13 US-09-969-984-46 Sequence 46, Appl
17 40.6 10.5 1092 17 US-10-437-963-97552 Sequence 97552, A
18 40 10.4 852 15 US-10-156-761-4858 Sequence 4858, Ap
c 19 40 10.4 9025608 15 US-10-156-761-1 Sequence 1, Appl1
20 39.8 10.3 1278 13 US-10-282-122A-11431 Sequence 11431, A
21 39.4 10.2 455 10 US-09-918-395-21800 Sequence 21800, A
22 39.4 10.2 474 10 US-09-918-395-21887 Sequence 21887, A
23 39.4 10.2 1354 9 US-09-764-898-99 Sequence 99, Appl
24 39.4 10.2 1354 16 US-10-264-049-523 Sequence 523, App
25 39.4 10.2 1371 9 US-09-764-898-26 Sequence 26, Appl
26 39.4 10.2 2040 16 US-10-104-047-1172 Sequence 1172, Ap
27 39.4 10.2 125401 17 US-10-203-295-35 Sequence 35, Appl
c 28 38.8 10.1 578 17 US-10-767-701-7923 Sequence 7923, Ap
29 38.8 10.1 1131 13 US-10-282-122A-13198 Sequence 13198, A
c 30 38.8 10.1 1428 17 US-10-437-963-78150 Sequence 78150, A
31 38.8 10.1 1479 16 US-10-260-238-1246 Sequence 1246, Ap
32 38.8 10.1 2851 13 US-10-425-114-31562 Sequence 31562, A
c 33 38.6 10.0 307 9 US-09-867-701-2574 Sequence 2574, Ap
34 38.6 10.0 1140 17 US-10-437-963-97551 Sequence 97551, A
35 38.6 10.0 21185 17 US-10-159-257A-2 Sequence 2, Appli
36 38.6 10.0 63158 15 US-10-252-198-1 Sequence 1, Appli
37 38 9.9 1500 13 US-10-412-699B-1133 Sequence 1133, Ap
38 38 9.9 1500 16 US-10-374-780A-680 Sequence 680, App
39 37.4 9.7 1287 15 US-10-156-761-3618 Sequence 3618, Ap
40 37.2 9.7 1200 9 US-09-909-320-22 Sequence 22, Appl
41 37.2 9.7 1200 9 US-09-909-088B-22 Sequence 22, Appl
42 37.2 9.7 1200 9 US-09-905-231A-22 Sequence 22, Appl
43 37.2 9.7 1200 9 US-09-902-853-22 Sequence 22, Appl
44 37.2 9.7 1200 9 US-09-907-824-22 Sequence 22, Appl
45 37.2 9.7 1200 9 US-09-907-841-22 Sequence 22, Appl

ALIGNMENTS

RESULT 1

US-10-077-130-3
; Sequence 3, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7893
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-077-130-3

Query Match 100.0%; Score 385; DB 14; Length 7893;
Best Local Similarity 100.0%; Pred. No. 5.8e-102;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCAGGCGAGATCTTTGACATCTAGCTGGTCAACGCTGACTACCTGCCCTAGGGGCTGA 60
777 CCAGGCGAGATCTTTGACATCTAGCTGGTCAACGCTGACTACCTGCCCTAGGGGCTGA 836
Db 61 GGAGGATGCATCAGCTCGGGAAGCCAGATGTGGAGGTCTTGGATGCAGCCACCC 120
837 GGAGGATGCATCAGCTCGGGAAGCCAGATGTGGAGGTCTTGGATGCAGCCACCC 896
QY 121 ACTGCGCTGGCTGTTCGCCACCAAGCCCAAGTCCAGCCCTCAAGGCTGGGT 180
897 ACTGCGCTGGCTGTTCGCCACCAAGCCCAAGTCCAGCCCTCAAGGCTGGGT 956

QY 181 GTCACAGCCTACCTGGACAGAGGCTCAAGCTGTACCTGTACCTGAGTGGGGGCCCGCTGAGGC 240
DB 957 GTCACAGCCTACCTGGACAGAGGCTCAAGCTGTACCTGTACCTGAGTGGGGGCCCGCTGAGGC 1016
QY 241 CCCTGAGTTCCTGGGGAGGCTGTGTCTGAAGACGAATACAAGCAAGGCTGAGCTCTGT 300
DB 1017 CCCTGAGTTCCTGGGGAGGCTGTGTCTGAAGACGAATACAAGCAAGGCTGAGCTCTGT 1076
QY 301 GATCAGAGCTGTGAGTTCGTGACAGGCTTCGTGGAGAGCTGCAGTTCCTGCAGAG 360
DB 1077 GATCAGAGCTGTGAGTTCGTGACAGGCTTCGTGGAGAGCTGCAGTTCCTGCAGAG 1136
QY 361 CCACCACTGCAGCACCTGGAGCGC 385
DB 1137 CCACCACTGCAGCACCTGGAGCGC 1161

RESULT 2

US-10-307-019-5
; Sequence 5, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenlin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SCIOS, INC.
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DV1
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 7928
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (60)...(7847)
US-10-307-019-5

Query Match 100.0%; Score 385; DB 15; Length 7928;
Best Local Similarity 100.0%; Pred. No. 5.8e-102;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGGGGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACTCTGCCCTAGGGGCTGA 60
DB 734 CCGAGGGGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACTCTGCCCTAGGGGCTGA 793
QY 61 GCAGGATGCCATCAGCTGCGGAAAGCCAGTATGTGGAGTCTCTGGATGCAGCCACCC 120
DB 794 GCAGGATGCCATCAGCTGCGGAAAGCCAGTATGTGGAGTCTCTGGATGCAGCCACCC 853
QY 121 ACTGCGTGTGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGGCTGGGT 180
DB 854 ACTGCGTGTGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGGCTGGGT 913
QY 181 GTACCAAGCCTA CTTGACAGAGGCTCAAGCTGTCTACCTGAGTGGGGGCCCGCTGAGGC 240
DB 914 GTACCAAGCCTA CTTGACAGAGGCTCAAGCTGTCTACCTGAGTGGGGGCCCGCTGAGGC 973
QY 241 CCCTGAGTTCCTGGGGAGGCTGTGTCTGAAGACGAATACAAGCAAGGCTGAGCTCTGT 300
DB 974 CCCTGAGTTCCTGGGGAGGCTGTGTCTGAAGACGAATACAAGCAAGGCTGAGCTCTGT 1033
QY 301 GATCAGAGCTGTGAGTTCGTGAGCAGGCTTCGTGGAGAGCTGCAGTTCCTGCAGAG 360
DB 1034 GATCAGAGCTGTGAGTTCGTGAGCAGGCTTCGTGGAGAGCTGCAGTTCCTGCAGAG 1093

QY 361 CCACCACTGCAGCACCTGGAGCGC 385
DB 1094 CCACCACTGCAGCACCTGGAGCGC 1118

RESULT 3

US-10-077-130-1
; Sequence 1, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; FILE REFERENCE: MPI2001-047PIRCPI(W)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 8106
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5' UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(7964)
; NAME/KEY: 3' UTR
; LOCATION: (7965)...(8106)
US-10-077-130-1

Query Match 100.0%; Score 385; DB 14; Length 8106;
Best Local Similarity 100.0%; Pred. No. 5.8e-102;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGGGGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACTCTGCCCTAGGGGCTGA 60
DB 848 CCGAGGGGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACTCTGCCCTAGGGGCTGA 907
QY 61 GCAGGATGCCATCAGCTGCGGAAAGCCAGTATGTGGAGTCTCTGGATGCAGCCACCC 120
DB 908 GCAGGATGCCATCAGCTGCGGAAAGCCAGTATGTGGAGTCTCTGGATGCAGCCACCC 967
QY 121 ACTGCGTGTGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGGCTGGGT 180
DB 968 ACTGCGTGTGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGGCTGGGT 1027
QY 181 GTCACCAAGCTTACCTGGACAGAGGCTCAAGCTGTCTACCTGAGTGGGGGCCCGCTGAGGC 240
DB 1028 GTCACCAAGCTTACCTGGACAGAGGCTCAAGCTGTCTACCTGAGTGGGGGCCCGCTGAGGC 1087
QY 241 CCCTGAGTTCCTGGGGAGGCTGTGTCTGAAGACGAATACAAGCAAGGCTGAGCTCTGT 300
DB 1088 CCCTGAGTTCCTGGGGAGGCTGTGTCTGAAGACGAATACAAGCAAGGCTGAGCTCTGT 1147
QY 301 GATCAGAGCTGTGAGTTCGTGAGCAGGCTTCGTGGAGAGCTGCAGTTCCTGCAGAG 360
DB 1148 GATCAGAGCTGTGAGTTCGTGAGCAGGCTTCGTGGAGAGCTGCAGTTCCTGCAGAG 1207
QY 361 CCACCACTGCAGCACCTGGAGCGC 385
DB 1208 CCACCACTGCAGCACCTGGAGCGC 1232

RESULT 4

US-10-077-130-6
; Sequence 6, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana

```
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6

Query Match          100.0%; Score 385; DB 14; Length 23907;
Best Local Similarity 100.0%; Pred. No. 6.5e-102;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTTACCTGCCCTTAGGGGCTGA 60
Db 16791 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTTACCTGCCCTTAGGGGCTGA 16850

QY 61 GCAGATGCCATCAGCTGCGGGAAGGCCAGTATGTGAGGTCTCTGATGAGCCCAACC 120
Db 16851 GCAGATGCCATCAGCTGCGGGAAGGCCAGTATGTGAGGTCTCTGATGAGCCCAACC 16910

QY 121 ACTGCGTGGCTGTTCGACCAACAGCCCAAGTCCAGCCCTCAGCGAGGGCTGGT 180
Db 16911 ACTGCGTGGCTGTTCGACCAACAGCCCAAGTCCAGCCCTCAGCGAGGGCTGGT 16970

QY 181 GTCACAGCTTACCTGGACAGAGGCTCAAGCTGTCAAGTGTGAGGAGGCTGAGCTCTGT 240
Db 16971 GTCACAGCTTACCTGGACAGAGGCTCAAGCTGTCAAGTGTGAGGAGGCTGAGCTCTGT 17030

QY 241 CCCTGAGTTCCTGGGAGGCTGTCTGAGAGCAATACAAAGCAAGGCTGAGCTCTGT 300
Db 17031 CCCTGAGTTCCTGGGAGGCTGTCTGAGAGCAATACAAAGCAAGGCTGAGCTCTGT 17090

QY 301 GATCAGGAGCTGTGAGTCTTGAGCAGGCGCTTCTGAGAGAGCTGAGTTCCTGAGAG 360
Db 17091 GATCAGGAGCTGTGAGTCTTGAGCAGGCGCTTCTGAGAGAGCTGAGTTCCTGAGAG 17150

QY 361 CCACACCTTGACGACCTGGAGCGC 385
Db 17151 CCACACCTTGACGACCTGGAGCGC 17175
```

```
RESULT 5
US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
```

```
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
US-10-077-130-4

Query Match          100.0%; Score 385; DB 14; Length 24120;
Best Local Similarity 100.0%; Pred. No. 6.5e-102;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTTACCTGCCCTTAGGGGCTGA 60
Db 16862 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTTACCTGCCCTTAGGGGCTGA 16921

QY 61 GCAGATGCCATCAGCTGCGGGAAGGCCAGTATGTGAGGTCTCTGATGAGCCCAACC 120
Db 16922 GCAGATGCCATCAGCTGCGGGAAGGCCAGTATGTGAGGTCTCTGATGAGCCCAACC 16981

QY 121 ACTGCGTGGCTGTTCGACCAACAGCCCAAGTCCAGCCCTCAGCGAGGGCTGGT 180
Db 16982 ACTGCGTGGCTGTTCGACCAACAGCCCAAGTCCAGCCCTCAGCGAGGGCTGGT 17041

QY 181 GTCACAGCTTACCTGGACAGAGGCTCAAGCTGTCAAGTGTGAGGAGGCTGAGCTCTGT 240
Db 17042 GTCACAGCTTACCTGGACAGAGGCTCAAGCTGTGAGGAGGCTGAGCTCTGT 17101

QY 241 CCCTGAGTTCCTGGGAGGCTGTCTGAGAGCAATACAAAGCAAGGCTGAGCTCTGT 300
Db 17102 CCCTGAGTTCCTGGGAGGCTGTCTGAGAGCAATACAAAGCAAGGCTGAGCTCTGT 17161

QY 301 GATCAGGAGCTGTGAGTCTTGAGCAGGCGCTTCTGAGAGAGCTGAGTTCCTGAGAG 360
Db 17162 GATCAGGAGCTGTGAGTCTTGAGCAGGCGCTTCTGAGAGAGCTGAGTTCCTGAGAG 17221

QY 361 CCACACCTTGACGACCTGGAGCGC 385
Db 17222 CCACACCTTGACGACCTGGAGCGC 17246
```

```
RESULT 6
US-09-864-408A-1271
; Sequence 1271, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Enc
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1271
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1)
; OTHER INFORMATION: Wherein n may be a, c, g or t
US-09-864-408A-1271
```

```
Query Match          97.9%; Score 377; DB 11; Length 628;
Best Local Similarity 100.0%; Pred. No. 9.5e-100;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AGATCTTTGACATCTACGTGTGTCACCGCTGACTTACCTGCCCTTAGGGGCTGAGCAGATG 68
Db 2 AGATCTTTGACATCTACGTGTGTCACCGCTGACTTACCTGCCCTTAGGGGCTGAGCAGATG 61

QY 69 CCATCAGCTGCGGGAAGGCCAGTATGTGAGGTCTCTGATGAGCCCAACCCTGCGCT 128
```

Db 62 CCATCACGCTCGGGAGGCCAGTATGTGGAGTCTTGATGACGCCACCACCTGGCT 121
 Qy 129 GGCCTGTGCGCACCAAGCCCAAGTCCAGCCCTCAGCGCAGGCTGGGTGTCACAG 188
 Db 122 GGCCTGTGCGCACCAAGCCCAAGTCCAGCCCTCAGCGCAGGCTGGGTGTCACAG 181
 Qy 189 CCTACTGGACAGGAGGCTCAAGCTGTACCTGAGTGGGGGCGCTGAGGCCCTGAGT 248
 Db 182 CCTACTGGACAGGAGGCTCAAGCTGTACCTGAGTGGGGGCGCTGAGGCCCTGAGT 241
 Qy 249 TCCCTGGGAGGCTGTGTCTGAAGACGAATACAAAGCAAGGCTGAGCTGTGTATCCAGG 308
 Db 242 TCCCTGGGAGGCTGTGTCTGAAGACGAATACAAAGCAAGGCTGAGCTGTGTATCCAGG 301
 Qy 309 AGCTGCTGAGTTCTGAGCAGCCCTTCGTGGAGGAGCTGCAGTTCTCTGACAGGCCACCACC 368
 Db 302 AGCTGCTGAGTTCTGAGCAGCCCTTCGTGGAGGAGCTGCAGTTCTCTGACAGGCCACCACC 361
 Qy 369 TGCAGCACTGGAGGC 385
 Db 362 TGCAGCACTGGAGGC 378

RESULT 7

US-10-156-761-393
 ; Sequence 393, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 393
 ; LENGTH: 1290
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1290)
 US-10-156-761-393

Query Match 12.3%; Score 47.4; DB 15; Length 1290;
 Best Local Similarity 49.0%; Pred. No. 0.00084;
 Matches 126; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
 Qy 18 ACATCTACGTGGTCCACCGCTGACTACCTGCCCCCTAGGGGTGAGCAGGATGCCATCACGC 77
 Db 305 ACATCGACCTCGACGACGCGACGACTGCGGTGCGGCAGTTACGTCGTCTACTCTCAAGG 364
 Qy 78 TGGGGAGGCCAGTATGTGGAGTCTTGATGACGCCACCACCTGCGGTGGCTGTGCTC 137
 Db 365 TCTGGGAACGCCACGTCACCGAGTGCAGGACCCCTCGTCTCCACGAACCCGCGCTCGGCC 424
 Qy 138 GCACCAAGCCCAAGTCCAGCCCTCAGCGCAGGCTGGGTGTACACGAGCTTACCTGG 197
 Db 425 TGCACCAAGCCGACTCGCGAGCGCGCAGGTGTGTGGCAGGTCCGGGCACTGCCG 484
 Qy 198 ACAGGAGGCTCAAGTGTACCTGAGTGGGGGCGCGTGAAGGCCCTGAGTTCCTGGGG 257
 Db 485 TCAAGAGCGGCGAAGAACCCCGTGCACCGGATCGCCCGGAGCAGCAGCCGACCGGGCA 544

Qy 258 AGGCTGTGTCTGAAGAC 274
 Db 545 CGGGCGTCTCGACGTC 561

RESULT 8

US-10-156-761-1
 ; Sequence 1, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Query Match 12.3%; Score 47.4; DB 15; Length 9025608;
 Best Local Similarity 49.0%; Pred. No. 0.0016;
 Matches 126; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
 Qy 18 ACATCTACGTGGTCCACCGCTGACTACCTGCCCCCTAGGGGTGAGCAGGATGCCATCACGC 77
 Db 471875 ACATCGACCTCGACGACGCGCAGCTGCGGTGCGGCAGTTACGTCGTCTACTCTCAAGG 471934
 Qy 78 TCGGGAGGCCAGTATGTGGAGTCTTGATGACGCCACCACCTGCGGTGGCTGTGCTC 137
 Db 471935 TCTGGGAACGCCACGTCACCGAGTGCAGGACCCCTCGTCTCCAGAACCCGCGCTCGGCC 471994
 Qy 138 GCACCAAGCCCAAGTCCAGCCCTCAGCGCAGGCTGGGTGTACACGAGCTTACCTGG 197
 Db 471995 TGCACCAAGCCGACTCGCGAGCGCGCAGGTGTGTGGCAGGTCCGGGCACTGCCG 472054
 Qy 198 ACAGGAGGCTCAAGTGTACCTGAGTGGGGGCGCGTGAAGGCCCTGAGTTCCTGGGG 257
 Db 472055 TCAAGAGCGGCGAAGAACCCCGTGCACCGGATCGCCCGGAGCAGCAGCCGACCGGGCA 472114
 Qy 258 AGGCTGTGTCTGAAGAC 274
 Db 472115 CGGGCGTCTCGACGTC 472131

RESULT 9

US-10-342-887-1266
 ; Sequence 1266, Application US/10342887
 ; Publication No. US20040058340A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dai, Hongyue
 ; APPLICANT: He, Yuding
 ; APPLICANT: Linsley, Peter S.
 ; APPLICANT: Mao, Mao
 ; APPLICANT: Roberts, Christopher J.
 ; APPLICANT: Van 't Veer, Laura Johanna
 ; APPLICANT: Van de Vijver, Marc J.

APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; PRIOR FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1266
; LENGTH: 4180
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-1266

Query Match 11.1%; Score 42.6; DB 13; Length 4180;
Best Local Similarity 48.5%; Pred. No. 0.018;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGGCTGGGTGTACACAGCCTACTCGACAGGA 203
Db |||||
313 AGCCGAGGAGCGAGCGGTTCAGGACAGGGGTGGATGGCCGCCCCACACAGTGGAGG 372
QY 204 GGCTCAAGCTGTACCTAGTGGGGGCGCTGAGGCGCCCTGTAGTTCCCTGGGGAGGCTG 263
Db |||||
373 AGCTGAAGGCGGCTGTGAGCGCGGCGAGCTGGAGGGCGGCGCGCTGTGCGGCTGG 432
QY 264 TGTCTGAAGACGAATACAGCAAGCTGAGCTGTGTATCCAGGAGCTCTGAGTTCTG 323
Db |||||
433 AGCGGAGCTGGCG 492
QY 324 AGCAGGCGCTTCGTGGAGGAGCTGCAAGTTCCTGACAGAGCCACACCTGAGCAGCCTGGAGC 383
Db |||||
493 GCCAGAGCAAGGTGGAGGCGCTGTACGAGCTGCTGCGGAGCCAGAGTGTGGGCGTGTGC 552
QY 384 G 384
Db 553 G 553

RESULT 10
US-10-172-118-1266
; Sequence 1266, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1266
; LENGTH: 4180
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM 006291
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-1266

Query Match 11.1%; Score 42.6; DB 13; Length 4180;

Best Local Similarity 48.5%; Pred. No. 0.018;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGGCTGGGTGTACACAGCCTACTCGACAGGA 203
Db |||||
313 AGCCGAGGAGCGAGCGGTTCAGGACAGGGGTGGATGGCCGCCCCACACAGTGGAGG 372
QY 204 GGCTCAAGCTGTACCTAGTGGGGGCGCTGAGGCGCCCTGTAGTTCCCTGGGGAGGCTG 263
Db |||||
373 AGCTGAAGGCGGCTGTGAGCGCGGCGAGCTGGAGGGCGGCGCGCTGTGCGGCTGG 432
QY 264 TGTCTGAAGACGAATACAGCAAGCTGAGCTGTGTATCCAGGAGCTCTGAGTTCTG 323
Db |||||
433 AGCGGAGCTGGCG 492
QY 324 AGCAGGCGCTTCGTGGAGGAGCTGCAAGTTCCTGACAGAGCCACACCTGAGCAGCCTGGAGC 383
Db |||||
493 GCCAGAGCAAGGTGGAGGCGCTGTACGAGCTGCTGCGGAGCCAGAGTGTGGGCGTGTGC 552
QY 384 G 384
Db 553 G 553

RESULT 11
US-10-439-388-10
; Sequence 10, Application US/10439388
; Publication No. US20030228617A1
; GENERAL INFORMATION:
; APPLICANT: Aune, Thomas M
; APPLICANT: Olsen, Nancy J
; TITLE OF INVENTION: Method for Predicting Autoimmune Disease
; FILE REFERENCE: 1242/68
; CURRENT APPLICATION NUMBER: US/10/439,388
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: US 60/381,055
; PRIOR FILING DATE: 2002-05-16
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 4180
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-439-388-10

Query Match 11.1%; Score 42.6; DB 16; Length 4180;
Best Local Similarity 48.5%; Pred. No. 0.018;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGGCTGGGTGTACACAGCCTACTCGACAGGA 203
Db |||||
313 AGCCGAGGAGCGAGCGGTTCAGGACAGGGGTGGATGGCCGCCCCACACAGTGGAGG 372
QY 204 GGCTCAAGCTGTACCTAGTGGGGGCGCTGAGGCGCCCTGAGTTCCCTGGGGAGGCTG 263
Db |||||
373 AGCTGAAGGCGGCTGTGAGCGCGGCGAGCTGGAGGGCGGCGCGCTGTGCGGCTGG 432
QY 264 TGTCTGAAGACGAATACAGCAAGCTGAGCTGTGTATCCAGGAGCTCTGAGTTCTG 323
Db |||||
433 AGCGGAGCTGGCG 492
QY 324 AGCAGGCGCTTCGTGGAGGAGCTGCAAGTTCCTGACAGAGCCACACCTGAGCAGCCTGGAGC 383
Db |||||
493 GCCAGAGCAAGGTGGAGGCGCTGTACGAGCTGCTGCGGAGCCAGAGTGTGGGCGTGTGC 552
QY 384 G 384
Db 553 G 553

RESULT 12
US-10-776-827-93
; Sequence 93, Application US/10776827

Publication No. US20040132086A1
GENERAL INFORMATION:
APPLICANT: Horwitz, Kathryn
APPLICANT: Richer, Jennifer
TITLE OF INVENTION: Progesterone Receptor-Regulated Gene Expression and Methods Related
FILE REFERENCE: 2848-33
CURRENT APPLICATION NUMBER: US/10/776,827
CURRENT FILING DATE: 2004-02-10
PRIOR APPLICATION NUMBER: US/09/814,915
PRIOR FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: 60/214,870
PRIOR FILING DATE: 2000-06-28
NUMBER OF SEQ ID NOS: 108
SOFTWARE: PatentIn version 3.1
SEQ ID NO 93
LENGTH: 4180
TYPE: DNA
ORGANISM: Homo sapiens
US-10-776-827-93

Query Match 11.1%; Score 42.6; DB 17; Length 4180;
Best Local Similarity 48.5%; Pred. No. 0.018;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTACACAGCCTACTCGACAGGA 203
DB 313 AGCCGAGGAGCGAGCGGTTCAGGAGGCTGGATGGCCCGCCCGCCACAGTGGAGG 372
QY 204 GGCTCAAGCTGTCACTGAGTGGGGGCGGCTGAGGCGCCCTGAGTTCCTCGGGAGGCTG 263
DB 373 AGCTGAAGCGGCTGTGAGCGGCGGAGCTGGAGGCGGCGCGCTGCTGGCGCTGG 432
QY 264 TGTCTGAAGACGATACAGCAGGCTGAGCTGTGTATCCAGGAGCTGCTGAGTTCTG 323
DB 433 AGCGGAGCTGGCGGCGGCGGCGGCGGCTGTGAGCGAGGAGGAGCTGGTGGCGG 492
QY 324 AGCAGGCTTGTGAGGAGCTGAGTTCCTGTGAGAGCCACACCTGACAGCCTGAGG 383
DB 493 GCCAGAGCAGGTGGAGGCGCTGTACGAGCTGTGCGGACAGGCTGTGGCGTGTCTG 552
QY 384 G 384
DB 553 G 553

RESULT 13
US-09-778-927A-16
Sequence 16, Application US/09778927A
Patent No. US2002068342A1
GENERAL INFORMATION:
APPLICANT: KHOSRAVI, Rami et al.
TITLE OF INVENTION: NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES AND NOVEL
FILE REFERENCE: 2786-0160P
CURRENT APPLICATION NUMBER: US/09/778,927A
CURRENT FILING DATE: 2001-02-08
PRIOR APPLICATION NUMBER: IL 134453
PRIOR FILING DATE: 2000-02-09
PRIOR APPLICATION NUMBER: IL135341
PRIOR FILING DATE: 2000-03-29
NUMBER OF SEQ ID NOS: 81
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 16
LENGTH: 4282
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(4282)
OTHER INFORMATION: n = a, c, g, t any unknown or other
US-09-778-927A-16

Query Match 11.1%; Score 42.6; DB 9; Length 4282;
Best Local Similarity 48.5%; Pred. No. 0.018;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTACACAGCCTACTCGACAGGA 203
DB 526 AGCCGAGGAGCGAGCGGTTCAGGAGGCTGGATGGCCCGCCCGCCACAGTGGAGG 585
QY 204 GGCTCAAGCTGTCACTGAGTGGGGGCGGCTGAGGCGCCCTGAGTTCCTCGGGAGGCTG 263
DB 585 AGCTGAAGCGGCTGTGAGCGGCGGAGCTGGAGCGGCGGCGGCTGCTGGCGCTGG 645
QY 264 TGTCTGAAGACGATACAGCAGGCTGAGCTGTGTATCCAGGAGCTGCTGAGTTCTG 323
DB 646 AGCGGAGCTGGCGGCGGCGGCGGCGGCTGTGAGCGAGGAGGAGCTGGTGGCGG 705
QY 324 AGCAGGCTTGTGAGGAGCTGAGTTCCTGTGAGAGCCACACCTGACAGCCTGAGG 383
DB 706 GCCAGAGCAGGTGGAGGCGCTGTACGAGCTGTGCGGACAGGCTGTGGCGTGTCTG 765
QY 384 G 384
DB 766 G 766

RESULT 14
US-10-044-090-668
Sequence 668, Application US/10044090
Publication No. US20020137081A1
GENERAL INFORMATION:
APPLICANT: Olga Bandman
TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
FILE REFERENCE: PA-0028 US
CURRENT APPLICATION NUMBER: US/10/044,090
CURRENT FILING DATE: 2002-01-09
NUMBER OF SEQ ID NOS: 850
SOFTWARE: PERL Program
SEQ ID NO 668
LENGTH: 4366
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020137081A1 319136.21
LOCATION: 3345-3660
OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-668

Query Match 11.1%; Score 42.6; DB 14; Length 4366;
Best Local Similarity 48.5%; Pred. No. 0.018;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTACACAGCCTACTCGACAGGA 203
DB 412 AGCCGAGGAGCGAGCGGTTCAGGAGGCTGGATGGCCCGCCCGCCACAGTGGAGG 471
QY 204 GGCTCAAGCTGTCACTGAGTGGGGGCGGCTGAGGCGCCCTGAGTTCCTCGGGAGGCTG 263
DB 472 AGCTGAAGCGGCTGTGAGCGGCGGAGCTGGAGGCGGCGGCGGCTGCTGGCGCTGG 531
QY 264 TGTCTGAAGACGATACAGCAGGCTGAGCTGTGTATCCAGGAGCTGCTGAGTTCTG 323
DB 532 AGCGGAGCTGGCGGCGGCGGCGGCGGCTGTGAGCGAGGAGGAGCTGGTGGCGG 591
QY 324 AGCAGGCTTGTGAGGAGCTGAGTTCCTGTGAGAGCCACACCTGACAGCCTGAGG 383
DB 592 GCCAGAGCAGGTGGAGGCGCTGTACGAGCTGTGCGGACAGGCTGTGGCGTGTCTG 651
QY 384 G 384
DB 652 G 652

RESULT 15
US-09-965-528-46
; Sequence 46, Application US/09965528
; Publication No. US20020187523A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: TANG, Y. Tom
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
; APPLICANT: BURFORD, Neil
; APPLICANT: BANDMAN, Olga
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LU, Dvung Aina M.
; APPLICANT: PATTERSON, Chandra
; TITLE OF INVENTION: EXTRACELLULAR SIGNALING MOLECULES
; FILE REFERENCE: PF-0701 USA
; CURRENT APPLICATION NUMBER: US/09/965,528
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/134,949
; PRIOR FILING DATE: 1999-05-19
; PRIOR APPLICATION NUMBER: 60/144,270
; PRIOR FILING DATE: 1999-07-15
; PRIOR APPLICATION NUMBER: 60/146,700
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/157,508
; PRIOR FILING DATE: 1999-10-04
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PEARL Program
; SEQ ID NO 46
; LENGTH: 1803
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020187523A1 843193CB1
US-09-965-528-46

Query Match 10.8%; Score 41.4; DB 9; Length 1803;
Best Local Similarity 47.5%; Pred. No. 0.037;
Matches 123; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
Qy 115 CCACCCACTGGCTGGCTTGTCCGACCAAGCCCAAGTCCAGCCCTCACGGCAGGG 174
Db 1501 CCAAGGACTGGCTGGGTGGCGGAGGGGAGCCAGATCCCCGAGGGAGGACCCCTGAGGG 1560
Qy 175 CTGGGTGTACCCAGCTTACCTGGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGGCGC 234
Db 1561 CCGCGAAGATCCGAGCCCTCCAGCTGGGAAGGGGAGGCGCGGTGCCCCAGGGGCGGTGGC 1620
Qy 235 TGAGGCCCTGTAGTTCCTTGGGGAGGCTGTGTCTGAACGAATACAAGGCAAGGCTGAG 294
Db 1621 ACAGTGCCCTTCCCGGACGGGTGGCAGGGCCCTGGAGAGGAACTGAGTGTCACTCCTGAT 1680
Qy 295 CTCTGTGATCCAGGAGCTGCTGAGTTCGTGACAGGCTTCGTGGAGGAGCTGCAGTTCCT 354
Db 1681 CTCAGGCCACCAAGCTCTGCGGCTCCAGCGGGCTCTGAGAGCCCGCTGAAGCCCGCTGAAGTCA 1740
Qy 355 GCAGAGCCACCACTGCAG 373
Db 1741 GCGACTTAAGGCTTGCGAG 1759

Search completed: September 20, 2004, 02:52:12
Job time : 214.275 secs

This Page Blank (uspto)

| | | | | | | |
|----|--------|------|-------|----|---------------------|--------------------|
| 16 | 4533 | 11.0 | 871 | 14 | US-10-307-019-7 | Sequence 7, Appli |
| 17 | 4529 | 11.0 | 871 | 16 | US-10-311-034-20 | Sequence 20, Appli |
| 18 | 3580.5 | 8.7 | 26926 | 9 | US-09-759-508B-2 | Sequence 2, Appli |
| 19 | 2629 | 6.4 | 4162 | 15 | US-10-023-634-92 | Sequence 92, Appli |
| 20 | 2489 | 6.0 | 6642 | 15 | US-10-369-493-5013 | Sequence 5013, Ap |
| 21 | 2447.5 | 5.9 | 548 | 14 | US-10-307-019-8 | Sequence 8, Appli |
| 22 | 2415.5 | 5.9 | 548 | 14 | US-10-307-019-8 | Sequence 9, Appli |
| 23 | 2324 | 5.6 | 3208 | 15 | US-10-210-130-38 | Sequence 38, Appli |
| 24 | 2319 | 5.6 | 3268 | 15 | US-10-379-381-2 | Sequence 2, Appli |
| 25 | 2270 | 5.5 | 3252 | 15 | US-10-210-130-36 | Sequence 36, Appli |
| 26 | 2203 | 5.3 | 3262 | 15 | US-10-379-381-4 | Sequence 4, Appli |
| 27 | 2135 | 5.2 | 2380 | 12 | US-10-333-314-18 | Sequence 78, Appli |
| 28 | 2102 | 5.1 | 5635 | 16 | US-10-451-168-78 | Sequence 18, Appli |
| 29 | 2085.5 | 5.1 | 3186 | 15 | US-10-210-130-34 | Sequence 34, Appli |
| 30 | 2085 | 5.1 | 416 | 15 | US-10-108-460A-4389 | Sequence 4389, Ap |
| 31 | 2072.5 | 5.0 | 5636 | 14 | US-10-032-189-128 | Sequence 128, App |
| 32 | 2072.5 | 5.0 | 5636 | 15 | US-10-120-801-72 | Sequence 72, Appli |
| 33 | 2072.5 | 5.0 | 5636 | 15 | US-10-023-634-93 | Sequence 93, Appli |
| 34 | 2072.5 | 5.0 | 5636 | 16 | US-10-408-765A-1895 | Sequence 1895, Ap |
| 35 | 2067 | 5.0 | 2231 | 15 | US-10-379-381-5 | Sequence 5, Appli |
| 36 | 2067 | 5.0 | 2242 | 16 | US-10-408-765A-793 | Sequence 793, App |
| 37 | 2048 | 5.0 | 5198 | 12 | US-10-346-863-34 | Sequence 34, Appli |
| 38 | 2048 | 5.0 | 5198 | 15 | US-10-120-801-75 | Sequence 75, Appli |
| 39 | 2048 | 5.0 | 5198 | 15 | US-10-369-493-6858 | Sequence 6858, Ap |
| 40 | 2048 | 5.0 | 5198 | 15 | US-10-369-493-6860 | Sequence 6860, Ap |
| 41 | 2045 | 5.0 | 390 | 12 | US-10-425-114-37530 | Sequence 37530, A |
| 42 | 2043.5 | 5.0 | 5175 | 15 | US-10-120-801-74 | Sequence 74, Appli |
| 43 | 2043.5 | 5.0 | 5175 | 15 | US-10-369-493-6859 | Sequence 6859, Ap |
| 44 | 2043.5 | 5.0 | 5175 | 15 | US-10-369-493-6861 | Sequence 6861, Ap |
| 45 | 1840 | 4.5 | 3931 | 15 | US-10-120-801-18 | Sequence 18, Appli |

ALIGNMENTS

RESULT 1

US-10-077-130-5
; Sequence 5, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Thereof
; FILE REFERENCE: MPI2001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 7968
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-077-130-5

Query Match 100.0%; Score 41273; DB 13; Length 7968;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 7968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| Qy | 1 | MDQFSGAPRLTEPKAFVSVGKDATLSQIVGNPTPOVSWKDDQPVTAGARFLAQ | 60 |
| Db | 1 | MDQFSGAPRLTEPKAFVSVGKDATLSQIVGNPTPOVSWKDDQPVTAGARFLAQ | 60 |
| Qy | 61 | DGDLRLTLILDALGDSQYVCRAINAIGEAFAAAGLVDAEAAACAEQAPHLRPTSIR | 120 |
| Db | 61 | DGDLRLTLILDALGDSQYVCRAINAIGEAFAAAGLVDAEAAACAEQAPHLRPTSIR | 120 |
| Qy | 121 | VREGSEATFCRVGSGPRPAVSWKDGRRIGCEPDGPRVVEELGEASALRTRAAAPRDGG | 180 |
| Db | 121 | VREGSEATFCRVGSGPRPAVSWKDGRRIGCEPDGPRVVEELGEASALRTRAAAPRDGG | 180 |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:36:00 ; Search time 480 Seconds
(without alignments)
5323.448 Million cell updates/sec

Title: US-10-077-130-5
Perfect score: 41273
Sequence: 1 MDQFSGAPRLTEPKAFV.....RNEKRRALLYKEHNLAAQVR 7968

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|---------|-------------|--------|-------|---------------------|
| 1 | 41273 | 100.0 | 7968 | 13 | US-10-077-130-5 |
| 2 | 32134 | 77.9 | 6620 | 15 | US-10-080-334-290 |
| 3 | 32134 | 77.9 | 6620 | 16 | US-10-408-765A-2291 |
| 4 | 13710 | 33.2 | 2630 | 13 | US-10-077-130-2 |
| 5 | 13528 | 32.8 | 2596 | 14 | US-10-307-019-6 |
| 6 | 10519.5 | 25.5 | 4691 | 15 | US-10-093-463-72 |
| 7 | 10506 | 25.5 | 4675 | 15 | US-10-093-463-74 |
| 8 | 9834.5 | 23.8 | 2328 | 16 | US-10-476-397-4 |
| 9 | 8423 | 20.4 | 1665 | 9 | US-09-858-664A-2 |
| 10 | 8423 | 20.4 | 1665 | 12 | US-10-697-263-2 |
| 11 | 8410 | 20.4 | 1655 | 12 | US-10-415-011-22 |
| 12 | 8407 | 20.4 | 1618 | 12 | US-10-182-243-46 |
| 13 | 8403 | 20.4 | 1610 | 14 | US-10-307-019-4 |
| 14 | 7073 | 17.1 | 1351 | 14 | US-10-307-019-1 |
| 15 | 6619.5 | 16.0 | 1596 | 16 | US-10-408-765A-992 |

181 TYEVRANPLGAASAAAALVDSADATASRPGTSTAALLAHLORREARAGAPASPP 240
181 TYEVRANPLGAASAAAALVDSADATASRPGTSTAALLAHLORREARAGAPASPP 240
241 STGTRTCTVTGEGHARLSCVITCEPKPETWKKDGQLVTEGRRHVYVEDAQENFVLKILF 300
241 STGTRTCTVTGEGHARLSCVITCEPKPETWKKDGQLVTEGRRHVYVEDAQENFVLKILF 300
301 CKOSDRGLYCTASNLVGYTSSVLVVRREPAPVFKRLODLVREKESATFLCEVPQPS 360
301 CKOSDRGLYCTASNLVGYTSSVLVVRREPAPVFKRLODLVREKESATFLCEVPQPS 360
361 TEAAWFKBEETRLWASAKYGIIEEGTERRLTVRNVASADDDAVYICETPEGSRTVAELAVOG 420
361 TEAAWFKBEETRLWASAKYGIIEEGTERRLTVRNVASADDDAVYICETPEGSRTVAELAVOG 420
421 NLLRKLPRKTAVRVGDPTAMFCVELAVPVPVHMLRNQOEVEVAGRVAIAEAGTRHTLTIS 480
421 NLLRKLPRKTAVRVGDPTAMFCVELAVPVPVHMLRNQOEVEVAGRVAIAEAGTRHTLTIS 480
481 OCCLEDVGOAFVAGDCQOTSTRCVSAPRKPLOPPVDVVKARMESSVILSWSPPHGE 540
481 OCCLEDVGOAFVAGDCQOTSTRCVSAPRKPLOPPVDVVKARMESSVILSWSPPHGE 540
541 RPTVIDGVLVEKKLGTITWIRCHEAEWVATPELTADVVAEENGFQFRVSALNSFGQSPY 600
541 RPTVIDGVLVEKKLGTITWIRCHEAEWVATPELTADVVAEENGFQFRVSALNSFGQSPY 600
601 LEFPGTHLAPKLVNTPKAVQAVEGGEVTFSDTLTVASAGWFLDGOALKASSVYIEH 660
601 LEFPGTHLAPKLVNTPKAVQAVEGGEVTFSDTLTVASAGWFLDGOALKASSVYIEH 660
661 CDRTRHTLTIREVPASLHGQALFXFVANGIESSIRMEVRAAPGLTANKPPAAAAREVLARL 720
661 CDRTRHTLTIREVPASLHGQALFXFVANGIESSIRMEVRAAPGLTANKPPAAAAREVLARL 720
721 HEEAQLLAELSDQAAAVTWIKOORTLSPGPKYEVQASAGRRVLLVDDVARDAGLYECVS 780
721 HEEAQLLAELSDQAAAVTWIKOORTLSPGPKYEVQASAGRRVLLVDDVARDAGLYECVS 780
781 RGGRIAYQLSVQGLARFLHKDMAGSCVDAVAGGPAOFECETSAHVHVHWYKDGMELGHS 840
781 RGGRIAYQLSVQGLARFLHKDMAGSCVDAVAGGPAOFECETSAHVHVHWYKDGMELGHS 840
841 GERFLQEDVGTNRHLVAATVTRDEGTYS CRVGEDSVDFRLRVSEPKVFAKEQLARRKL 900
841 GERFLQEDVGTNRHLVAATVTRDEGTYS CRVGEDSVDFRLRVSEPKVFAKEQLARRKL 900
901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMETGCTRRLVVQOAGQADAGEY 960
901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMETGCTRRLVVQOAGQADAGEY 960
961 SCAGGQRLSFHLDVKEPKVVFPAKQVVAHSEVOAEAGANATLSCEVAQAQAEVMMYKDGK 1020
961 SCAGGQRLSFHLDVKEPKVVFPAKQVVAHSEVOAEAGANATLSCEVAQAQAEVMMYKDGK 1020
1021 KLSSSLVHVYKACRRRLVVQOAGTKTDAGDYSCAARGQVSRFLHTBPXWMAFEQSV 1080
1021 KLSSSLVHVYKACRRRLVVQOAGTKTDAGDYSCAARGQVSRFLHTBPXWMAFEQSV 1080
1081 HNEVOAEAGASAMLSCEVAQAQTEVTWYKDGKLSSSSKVGMVEKGTCTRRLVLPQAGKAD 1140
1081 HNEVOAEAGASAMLSCEVAQAQTEVTWYKDGKLSSSSKVGMVEKGTCTRRLVLPQAGKAD 1140
1141 AGEYSCAGQORVSFHLHITTEPKGVFAKEQSVNEVOAEBAGTTAMLSCVAQOQTEVTWY 1200
1141 AGEYSCAGQORVSFHLHITTEPKGVFAKEQSVNEVOAEBAGTTAMLSCVAQOQTEVTWY 1200
1201 KDGKLSSSSKVNRMEVKGCTRRLVVQOAGKADAGEYSCAARGQVSRFLHTBPXWMAFEK 1260
1201 KDGKLSSSSKVNRMEVKGCTRRLVVQOAGKADAGEYSCAARGQVSRFLHTBPXWMAFEK 1260
1261 EQLVHNEVRTAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRIEAAAGCWRQLVVQQA 1320

1261 EQLVHNEVRTAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRIEAAAGCWRQLVVQQA 1320
1321 GOADAGEYTCBAGQORLSFHLVDSEPKAVFAKESOLAHKRVQAEAGATATLSCEVAQAQTE 1380
1321 GOADAGEYTCBAGQORLSFHLVDSEPKAVFAKESOLAHKRVQAEAGATATLSCEVAQAQTE 1380
1381 VTWYKDGKLSSSSKVNRMEAVGCTRRLVVQOACQADTGEYSCAARGQORLSFSDVDAEPKV 1440
1381 VTWYKDGKLSSSSKVNRMEAVGCTRRLVVQOACQADTGEYSCAARGQORLSFSDVDAEPKV 1440
1441 VFAKEQPVHREVOAQAQASTLSCEVAQAQTEVTWYKDGKLSFSSKVRMEAVGCTRRLV 1500
1441 VFAKEQPVHREVOAQAQASTLSCEVAQAQTEVTWYKDGKLSFSSKVRMEAVGCTRRLV 1500
1501 VQOAGQADAGEYSCAQSQRLSFHLHVAEPKAVFAKEQAPASREVQAEAGTATLSCEVAQ 1560
1501 VQOAGQADAGEYSCAQSQRLSFHLHVAEPKAVFAKEQAPASREVQAEAGTATLSCEVAQ 1560
1561 AQTEVTWYKDGKLSSSSKVNRMEAVGCTRRLVVQOAGQADAGEYSCKAGQORLSFHLHVA 1620
1561 AQTEVTWYKDGKLSSSSKVNRMEAVGCTRRLVVQOAGQADAGEYSCKAGQORLSFHLHVA 1620
1621 EPKVVFAKEQPAHREVOAQAQASTLSCEVAQAQTEVTWYKDGKLSSSSKVRMEAVGCT 1680
1621 EPKVVFAKEQPAHREVOAQAQASTLSCEVAQAQTEVTWYKDGKLSSSSKVRMEAVGCT 1680
1681 RRLVVQOAGQADAGEYSCAQSQRLSFHLHVAEPOISERPCCRRPLVVKHEHDIILTA 1740
1681 RRLVVQOAGQADAGEYSCAQSQRLSFHLHVAEPOISERPCCRRPLVVKHEHDIILTA 1740
1741 TLATPSAATVTLKDGVEIRRSKHETASQDTHTLTVHGAQVLDLSAIYSCRVAEGQDF 1800
1741 TLATPSAATVTLKDGVEIRRSKHETASQDTHTLTVHGAQVLDLSAIYSCRVAEGQDF 1800
1801 PVQVEEVAAKFCRLLEPVCGLGTTVLACELSPACAEVWVRGNTOPRVKXRFQWVAEG 1860
1801 PVQVEEVAAKFCRLLEPVCGLGTTVLACELSPACAEVWVRGNTOPRVKXRFQWVAEG 1860
1861 PVRSLTVLGLRAEDAGEYVCESEDDHTSAQITVSVPVVKFMSGLSTVVAEAGEATFQC 1920
1861 PVRSLTVLGLRAEDAGEYVCESEDDHTSAQITVSVPVVKFMSGLSTVVAEAGEATFQC 1920
1921 VVSPSDVAVVWFRGALLQPSKFAISQSGASHLSTISDLVLEDAQOITVVEAGASSAA 1980
1921 VVSPSDVAVVWFRGALLQPSKFAISQSGASHLSTISDLVLEDAQOITVVEAGASSAA 1980
1981 LRVEAPVLFKKLEPQTVSEERSVTLLEVELTRPPELRTWRNATALAPGNVIEHAEGA 2040
1981 LRVEAPVLFKKLEPQTVSEERSVTLLEVELTRPPELRTWRNATALAPGNVIEHAEGA 2040
2041 RHRLVNLVNGFADRGFGCETPDKTOAKLTVMRQVRLYRGICAVEARCGGTATHEVOL 2100
2041 RHRLVNLVNGFADRGFGCETPDKTOAKLTVMRQVRLYRGICAVEARCGGTATHEVOL 2100
2101 SHADVGSWTRDGLRFQOGPTCHLAVRGPMTLTLGLRPEDSGLMVFKAEGVHTSARLV 2160
2101 SHADVGSWTRDGLRFQOGPTCHLAVRGPMTLTLGLRPEDSGLMVFKAEGVHTSARLV 2160
2161 VTELVPFSRPLQDQVVTTEKEKVTLECELSRPNDVRLWKDGVLELRAGKTMAIAAQACR 2220
2161 VTELVPFSRPLQDQVVTTEKEKVTLECELSRPNDVRLWKDGVLELRAGKTMAIAAQACR 2220
2221 SLTIYRCFADQGVVCDADHQAQSSAVKVGQRTYTLIYRRVLAEDAGEICQFVAENAESR 2280
2221 SLTIYRCFADQGVVCDADHQAQSSAVKVGQRTYTLIYRRVLAEDAGEICQFVAENAESR 2280
2281 AQLRVKELPVLVLRPLRDKTAMEKXRGVLECEQVSRASAOVWFKGSOELQFGPKYELVSD 2340
2281 AQLRVKELPVLVLRPLRDKTAMEKXRGVLECEQVSRASAOVWFKGSOELQFGPKYELVSD 2340
2341 GLYRKLIISDVHAEDEDTYTCADGDKTSAOFFVEBOSITIVGLQDQVTVWEPAPAFEC 2400

Db 2341 CJYRKLIIISDVHAEDEDTYTCADGDKTSAQFVVEQSIITVIRGLQDVTVMPEAPAFEC 2400
Qy 2401 ETSIPSVRPPKWLJGKTVLQAGNGVLEQEBQGVHRLMLRTCTMTGPHVFTVCKSRSSA 2460
Db 2401 ETSIPSVRPPKWLJGKTVLQAGNGVLEQEBQGVHRLMLRTCTMTGPHVFTVCKSRSSA 2460
Qy 2461 RLIVSDIPVLTPLPKTGRELOSIVLSCDFRPAKAVQWYKDDTPLSPSEKFKMSLEG 2520
Db 2461 RLIVSDIPVLTPLPKTGRELOSIVLSCDFRPAKAVQWYKDDTPLSPSEKFKMSLEG 2520
Qy 2521 QMAELRILRLMPADAGVYRCQAGSAHSSTEVTVZREAVTVTGPLODAEATEEGWASFSCE 2580
Db 2521 QMAELRILRLMPADAGVYRCQAGSAHSSTEVTVZREAVTVTGPLODAEATEEGWASFSCE 2580
Qy 2581 LSHDEEVEWELNGMPLINDSFHEISHKGRHTLVLSKIOPADAGIVRASLKVSTSAKL 2640
Db 2581 LSHDEEVEWELNGMPLINDSFHEISHKGRHTLVLSKIOPADAGIVRASLKVSTSAKL 2640
Qy 2641 EVRVKPVVFLKALDDLSAEERGTALQCEVSDPEAHVVRKDXGVLQGPSDKYDLFHTAGT 2700
Db 2641 EVRVKPVVFLKALDDLSAEERGTALQCEVSDPEAHVVRKDXGVLQGPSDKYDLFHTAGT 2700
Qy 2701 RGLVVHDVSPDAGLYTCHVGSEBTRARVRVHDLVHGITKELKTMEVLEBESCSFECVLS 2760
Db 2701 RGLVVHDVSPDAGLYTCHVGSEBTRARVRVHDLVHGITKELKTMEVLEBESCSFECVLS 2760
Qy 2761 HESASDPAMWTVGGKTVGSSSRFOATROGRXYILVWREAPSADAGEVVFVSRGITSKASL 2820
Db 2761 HESASDPAMWTVGGKTVGSSSRFOATROGRXYILVWREAPSADAGEVVFVSRGITSKASL 2820
Qy 2821 IVRRERPAALIKPLEDQWVAPGEDVELRCELRSAGTPVHMLKDKAIRKSKYDVVCEGTM 2880
Db 2821 IVRRERPAALIKPLEDQWVAPGEDVELRCELRSAGTPVHMLKDKAIRKSKYDVVCEGTM 2880
Qy 2881 AMLVIRGASLDAGEYTCVEASKSTASLHVEEKANCFTBELTNLOVEEKGTAVFTCKTE 2940
Db 2881 AMLVIRGASLDAGEYTCVEASKSTASLHVEEKANCFTBELTNLOVEEKGTAVFTCKTE 2940
Qy 2941 HPAATVTVWRKGLLELRASGKHQPSQEGTLTLTTSALEKADSDTYTCDIGQAQRAQLLV 3000
Db 2941 HPAATVTVWRKGLLELRASGKHQPSQEGTLTLTTSALEKADSDTYTCDIGQAQRAQLLV 3000
Qy 3001 QGRRVHIIIEDLVVDVQGSATFRCSRIPANYBPVFWFLDKTPIHANELNEIDAQPGGY 3060
Db 3001 QGRRVHIIIEDLVVDVQGSATFRCSRIPANYBPVFWFLDKTPIHANELNEIDAQPGGY 3060
Qy 3061 HVLTLRQLALKDSGTIFYEAGDQASALRVTEKPSVFSRELTDATITEGEDTLVLCETS 3120
Db 3061 HVLTLRQLALKDSGTIFYEAGDQASALRVTEKPSVFSRELTDATITEGEDTLVLCETS 3120
Qy 3121 TCDIPMCWTKDGKTLRGARCOLSHEGHRALQITGATLQDSGRYKCEAGACSSIVRV 3180
Db 3121 TCDIPMCWTKDGKTLRGARCOLSHEGHRALQITGATLQDSGRYKCEAGACSSIVRV 3180
Qy 3181 HARPVRFQALKDLEVLGGATLRCVLSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLE 3240
Db 3181 HARPVRFQALKDLEVLGGATLRCVLSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLE 3240
Qy 3241 LVVRNLRPDQSGRYSCSGDQTTSATLTVTALPAQFIGLKRKATEGATATLRCELSKT 3300
Db 3241 LVVRNLRPDQSGRYSCSGDQTTSATLTVTALPAQFIGLKRKATEGATATLRCELSKT 3300
Qy 3301 APVEWRKSETLRDGRYCLRODGMCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPM 3360
Db 3301 APVEWRKSETLRDGRYCLRODGMCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPM 3360
Qy 3361 PAHFTRGRLRHOBESIGATATLRCELSKAAPVEWRKGRSLRQDGRHSLRQDGAUCELOIC 3420
Db 3361 PAHFTRGRLRHOBESIGATATLRCELSKAAPVEWRKGRSLRQDGRHSLRQDGAUCELOIC 3420
Qy 3421 GLAVADAGEYSCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLCELSKVAPVE 3480
Db 3421 GLAVADAGEYSCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLCELSKVAPVE 3480

Qy 3481 WRKGPENLRDGRYTLRQEGTRCELOICGLAMADAGEVLCVCGOERTSATLTIRALPARF 3540
Db 3481 WRKGPENLRDGRYTLRQEGTRCELOICGLAMADAGEVLCVCGOERTSATLTIRALPARF 3540
Qy 3541 IEDVKNOEAREGATAVLOCELNSAAPVEWRKSETLRDGRYSLRQDGTKCELOIRGLAM 3600
Db 3541 IEDVKNOEAREGATAVLOCELNSAAPVEWRKSETLRDGRYSLRQDGTKCELOIRGLAM 3600
Qy 3601 ADTGEYSCVCGOERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKNAPVEWK 3660
Db 3601 ADTGEYSCVCGOERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKNAPVEWK 3660
Qy 3661 HETLRDGRHSRLRQDGRARCELOIRGLVAEDAGEYLCMCKERTSAMLTVRAMPSFIEGL 3720
Db 3661 HETLRDGRHSRLRQDGRARCELOIRGLVAEDAGEYLCMCKERTSAMLTVRAMPSFIEGL 3720
Qy 3721 RNEEATEGATATLWCELSKAAAPVEWRKGHETTLDGRHSRLRQDGRARCELOIRGLVAD 3780
Db 3721 RNEEATEGATATLWCELSKAAAPVEWRKGHETTLDGRHSRLRQDGRARCELOIRGLVAD 3780
Qy 3781 EYSCVCGOERTSATLTVRALPARFTEDVKNOEAREGATAVLOCELNSKAAAPVEWRKSETL 3840
Db 3781 EYSCVCGOERTSATLTVRALPARFTEDVKNOEAREGATAVLOCELNSKAAAPVEWRKSETL 3840
Qy 3841 RGGDRYSLRQDGTTRCELOIRGLSVADTGEYSCVCGOERTSATLTVRAPQVFPREPLOSLQ 3900
Db 3841 RGGDRYSLRQDGTTRCELOIRGLSVADTGEYSCVCGOERTSATLTVRAPQVFPREPLOSLQ 3900
Qy 3901 ABEGSTATLQCELSSEPTATVWMSKGLQLOQANGREPRELOGCTAEILVLODQREDTGEVT 3960
Db 3901 ABEGSTATLQCELSSEPTATVWMSKGLQLOQANGREPRELOGCTAEILVLODQREDTGEVT 3960
Qy 3961 CTCGSQATSATLTVAAPVRFLELQHQEVDEGGTAHLCCELSRAGASVEMWKGSLQLPF 4020
Db 3961 CTCGSQATSATLTVAAPVRFLELQHQEVDEGGTAHLCCELSRAGASVEMWKGSLQLPF 4020
Qy 4021 CARYQWQDGAELLVRGVEQEDAGDYTCDTGHTQSMASLSVVRPFRPFKTRLSQLEOE 4080
Db 4021 CARYQWQDGAELLVRGVEQEDAGDYTCDTGHTQSMASLSVVRPFRPFKTRLSQLEOE 4080
Qy 4081 TGDIALRCCQLSDAESGAVVQWLKEGVELHAGPKYEMRSQGATRELLIHQLEAKDTGEVA 4140
Db 4081 TGDIALRCCQLSDAESGAVVQWLKEGVELHAGPKYEMRSQGATRELLIHQLEAKDTGEVA 4140
Qy 4141 CVTGGOKTAASLRVTEPEVTIVRGLVDAEVTADEDFEFSCEVSRAGATGVQWCLQPLQ 4200
Db 4141 CVTGGOKTAASLRVTEPEVTIVRGLVDAEVTADEDFEFSCEVSRAGATGVQWCLQPLQ 4200
Qy 4201 SNEVTEVAVRDCGRHITLRLKGYTPEDAGTVSHLGNHASSAQLTVREAPVTILEPLQDVQ 4260
Db 4201 SNEVTEVAVRDCGRHITLRLKGYTPEDAGTVSHLGNHASSAQLTVREAPVTILEPLQDVQ 4260
Qy 4261 LSSGQDASFCRLSRASGQEARWALGCVPLQANEMNDITVEQGTLLHLLTHKVTLEDAGT 4320
Db 4261 LSSGQDASFCRLSRASGQEARWALGCVPLQANEMNDITVEQGTLLHLLTHKVTLEDAGT 4320
Qy 4321 VSHVGTCSSEAOQLKVTAKNTVVRGLNVEALEGGEALFECQLSQEVAHAHTWLLDDEPV 4380
Db 4321 VSHVGTCSSEAOQLKVTAKNTVVRGLNVEALEGGEALFECQLSQEVAHAHTWLLDDEPV 4380
Qy 4381 RTSENAEEVFFENGLRHLLLLKNLRPQDSCRVTFLAGDMVTSAFLTVRGRWLEILEPLKN 4440
Db 4381 RTSENAEEVFFENGLRHLLLLKNLRPQDSCRVTFLAGDMVTSAFLTVRGRWLEILEPLKN 4440
Qy 4441 AAVRAGAQAARFTCTLSEAVPVGEASWYINGAAVQPDSDSWTVTADOSHALLLRSQPHH 4500
Db 4441 AAVRAGAQAARFTCTLSEAVPVGEASWYINGAAVQPDSDSWTVTADOSHALLLRSQPHH 4500
Qy 4501 AGEVTFACRDVAVASARLTVLGLPDPPEDEAVVAHSSHTVTLSWAAPMSDGGGLCGYRVE 4560
Db 4501 AGEVTFACRDVAVASARLTVLGLPDPPEDEAVVAHSSHTVTLSWAAPMSDGGGLCGYRVE 4560

QY 4561 VKEGATGOWRLCHELVPGPEBCVWDLAPGETYRFRVAAGVGVGAGEPVHLPTQTVRLAEPP 4620
DB 4561 VKEGATGOWRLCHELVPGPEBCVWDLAPGETYRFRVAAGVGVGAGEPVHLPTQTVRLAEPP 4620
QY 4621 KPVPQPSAPSRQVAAGEDVSLLEVVAEAGEVIVHKGMERIOPGRFEEVVSQGRQML 4680
DB 4621 KPVPQPSAPSRQVAAGEDVSLLEVVAEAGEVIVHKGMERIOPGRFEEVVSQGRQML 4680
QY 4681 VIKGFTAEQGEYHCGLAAGSICPAATTFQVALSPASVDEAPQPSLPEAAQEGDLHLW 4740
DB 4681 VIKGFTAEQGEYHCGLAAGSICPAATTFQVALSPASVDEAPQPSLPEAAQEGDLHLW 4740
QY 4741 EALARKRMSREPTLDSISELPEEDGRSQRUPQAEVAPDLSSEGYSTADELARTGDADL 4800
DB 4741 EALARKRMSREPTLDSISELPEEDGRSQRUPQAEVAPDLSSEGYSTADELARTGDADL 4800
QY 4801 SHTSSDDESRACTPSLVTLKAGRPCTSPILASKVGAAPSVKPOQOQEPFLAARPLG 4860
DB 4801 SHTSSDDESRACTPSLVTLKAGRPCTSPILASKVGAAPSVKPOQOQEPFLAARPLG 4860
QY 4861 DLSKTOIGDPSMDKAAVKIQAAPFKYKVRKEMKQOEGMFMSHTFGDTEAQVGDALRLEC 4920
DB 4861 DLSKTOIGDPSMDKAAVKIQAAPFKYKVRKEMKQOEGMFMSHTFGDTEAQVGDALRLEC 4920
QY 4921 VASRADVRARWLKDGVELTDCGRHHIDOLGDTGCSLLIAGLDRADAGCYTCQVSNKFGQV 4980
DB 4921 VASRADVRARWLKDGVELTDCGRHHIDOLGDTGCSLLIAGLDRADAGCYTCQVSNKFGQV 4980
QY 4981 THSACVVVSGSESAESSGSELDDAFRRARLRHRLFRKSPAEVSEBELFTSADGPA 5040
DB 4981 THSACVVVSGSESAESSGSELDDAFRRARLRHRLFRKSPAEVSEBELFTSADGPA 5040
QY 5041 EPERPADQTVREDEHFCIRFEALTEARQAVTRFQEMFATLGIVGKILVQOQPRVEM 5100
DB 5041 EPERPADQTVREDEHFCIRFEALTEARQAVTRFQEMFATLGIVGKILVQOQPRVEM 5100
QY 5101 CISKETPAPVPPPEPLPSLLTSDAAPVLTTELQVQVQDGYVPSFDCVVTGQPMPSVRWF 5160
DB 5101 CISKETPAPVPPPEPLPSLLTSDAAPVLTTELQVQVQDGYVPSFDCVVTGQPMPSVRWF 5160
QY 5161 KDGKLEEDHVMINEQOQGHQIITAVVPADMGVYRCLEANSMGVSTKAEILVDLTS 5220
DB 5161 KDGKLEEDHVMINEQOQGHQIITAVVPADMGVYRCLEANSMGVSTKAEILVDLTS 5220
QY 5221 TDYTAADATESSSYFSAQGYLSREOEGTESITTDGQLPQVVEELRDQVAPGTRLAKF 5280
DB 5221 TDYTAADATESSSYFSAQGYLSREOEGTESITTDGQLPQVVEELRDQVAPGTRLAKF 5280
QY 5281 QLKVGYPAPPLYNFKGQOPLTASAHIRMTGKLLHLEIISVTRSDSGQYAYISNMG 5340
DB 5281 QLKVGYPAPPLYNFKGQOPLTASAHIRMTGKLLHLEIISVTRSDSGQYAYISNMG 5340
QY 5341 AAYSARLLVRGPEPEEKPADSVHEQLVPPMLERFTPKYKXGSSITFSVXVGEKRPV 5400
DB 5341 AAYSARLLVRGPEPEEKPADSVHEQLVPPMLERFTPKYKXGSSITFSVXVGEKRPV 5400
QY 5401 TVHMLREABEGVWIGPDPGYTVASAOHSLVLDVGRQHQGTTCIASNAAGALC 5460
DB 5401 TVHMLREABEGVWIGPDPGYTVASAOHSLVLDVGRQHQGTTCIASNAAGALC 5460
QY 5461 SASLHVSLPKVBEQEKVKEALISTFLOQTQATISAOGLTAGFADLGGQKEEPLAAKE 5520
DB 5461 SASLHVSLPKVBEQEKVKEALISTFLOQTQATISAOGLTAGFADLGGQKEEPLAAKE 5520
QY 5521 ALGHLSLAEVTEBEFLKLTQITEMVSAKITQAKLVQPGDSDDESKTPSAPRGRSR 5580
DB 5521 ALGHLSLAEVTEBEFLKLTQITEMVSAKITQAKLVQPGDSDDESKTPSAPRGRSR 5580
QY 5581 PSSSIQSSSESSEGDARGEIFDIYVVTADYLPGLAFQDDAITREGOYVEVLDAAPLW 5640
DB 5581 PSSSIQSSSESSEGDARGEIFDIYVVTADYLPGLAFQDDAITREGOYVEVLDAAPLW 5640
QY 5641 LVRTKTKSGSPRQGVSPAYLDRRLKLSPEWGAEEAFEPFGAVSEDEYKARLSSVIOE 5700

DB 5641 LVRTKTKSGSPRQGVSPAYLDRRLKLSPEWGAEEAFEPFGAVSEDEYKARLSSVIOE 5700
QY 5701 LLSSEQAFVEELQFLQSHHLQHLERCCHPVPIAVAGOKAVIPRNVRDIDGRFHSSFLQELQ 5760
DB 5701 LLSSEQAFVEELQFLQSHHLQHLERCCHPVPIAVAGOKAVIPRNVRDIDGRFHSSFLQELQ 5760
QY 5761 CDTDDVAMCFIKNOAAFEQYLFLVGRVQAEVVVSTAIOEFYKXAEALLAGDPSQ 5820
DB 5761 CDTDDVAMCFIKNOAAFEQYLFLVGRVQAEVVVSTAIOEFYKXAEALLAGDPSQ 5820
QY 5821 PPPPLQHYLEQVERVQYQALLKELIRNKARNRQNCALLQAYAVVSAALFORAENKLHV 5880
DB 5821 PPPPLQHYLEQVERVQYQALLKELIRNKARNRQNCALLQAYAVVSAALFORAENKLHV 5880
QY 5881 SLWENYPTGLEALGEPPIRQGHFIVMEGAPGARMKGNHNVFLFRNLHVICRPRDSRT 5940
DB 5881 SLWENYPTGLEALGEPPIRQGHFIVMEGAPGARMKGNHNVFLFRNLHVICRPRDSRT 5940
QY 5941 DTVSVYFRNMKLSIDLDNDQVEGDDRAFEVWQEREDSVKYLLOARTAILKSSWVKEIC 6000
DB 5941 DTVSVYFRNMKLSIDLDNDQVEGDDRAFEVWQEREDSVKYLLOARTAILKSSWVKEIC 6000
QY 6001 GICQORLALPVWRPPEBELADCTAELGETVKLACRVGTGTPKVISWYKDKGKAVQVDPHH 6060
DB 6001 GICQORLALPVWRPPEBELADCTAELGETVKLACRVGTGTPKVISWYKDKGKAVQVDPHH 6060
QY 6061 ILIIBPDGSCALILDSLTGVDSGQWCMCPAASAGNCSTLGLKILVQVPPFRVNVKVRASP 6120
DB 6061 ILIIBPDGSCALILDSLTGVDSGQWCMCPAASAGNCSTLGLKILVQVPPFRVNVKVRASP 6120
QY 6121 EGEDAQFTCTTEGAPYPOIRWYKDGALLTGNKFTLSEPRSGLLVLVIRAAKDEDGLY 6180
DB 6121 EGEDAQFTCTTEGAPYPOIRWYKDGALLTGNKFTLSEPRSGLLVLVIRAAKDEDGLY 6180
QY 6181 ECELNVLGSRASAEALRIQSPMLQAOCHREOLVAAVEDTTILERADQEVTSVKRLLG 6240
DB 6181 ECELNVLGSRASAEALRIQSPMLQAOCHREOLVAAVEDTTILERADQEVTSVKRLLG 6240
QY 6241 PKAPGSGDITGPGPCPRGAPALQETGSOQPPVTGTSAPAVPPRVPOLLHEGPEOPE 6300
DB 6241 PKAPGSGDITGPGPCPRGAPALQETGSOQPPVTGTSAPAVPPRVPOLLHEGPEOPE 6300
QY 6301 ATARAQEVTVPIRMEGAAMPAGAGTGELLVDVSHVVRVETTTQRTYTYQAIDTHTARPSP 6360
DB 6301 ATARAQEVTVPIRMEGAAMPAGAGTGELLVDVSHVVRVETTTQRTYTYQAIDTHTARPSP 6360
QY 6361 VTIEDVQAGTGGTAQFAEALIEGDPQPSVTWYKDSVOLVDSTLSQOQEGTYSVLARHA 6420
DB 6361 VTIEDVQAGTGGTAQFAEALIEGDPQPSVTWYKDSVOLVDSTLSQOQEGTYSVLARHA 6420
QY 6421 SKDAGVYTCLAQNTGGVLCMAELVLDGNDNEPDSQSHRKLHSFVEKXEEIGRGVFG 6480
DB 6421 SKDAGVYTCLAQNTGGVLCMAELVLDGNDNEPDSQSHRKLHSFVEKXEEIGRGVFG 6480
QY 6481 FVKRVQHGKXKILCAAKIPIPLSRTRAQAYBERDILAAALSHPLVTGLLDQETRTKILLI 6540
DB 6481 FVKRVQHGKXKILCAAKIPIPLSRTRAQAYBERDILAAALSHPLVTGLLDQETRTKILLI 6540
QY 6541 LELCSSELLDRLYKRVVTEAEVKVYIQQVLVEGLVHLSHGVHLHLDIKPSNIMLVHPAR 6600
DB 6541 LELCSSELLDRLYKRVVTEAEVKVYIQQVLVEGLVHLSHGVHLHLDIKPSNIMLVHPAR 6600
QY 6601 EDIKICDFGPAQNTIPAELOFSQYSGSPFVSPETIIQONPVSEASDIWAMGVISYLSITS 6660
DB 6601 EDIKICDFGPAQNTIPAELOFSQYSGSPFVSPETIIQONPVSEASDIWAMGVISYLSITS 6660
QY 6661 SPFAGESDRATLLNVLEGRVSWSPMAHLSEDAKDFIKATLQAPARSAACLSHPW 6720
DB 6661 SPFAGESDRATLLNVLEGRVSWSPMAHLSEDAKDFIKATLQAPARSAACLSHPW 6720
QY 6721 FLKGNVPAEEAHFINTKQLKLLARSRQWSLMGYSILVVRSTPELLRGPPDPSLGVAR 6780

Db 6721 FLKMPAAEEAHFINTKQLKFLARSRWQSLMSYKSLVNRISPEILLRGPDPSPSLGVAR 6780
Qy 6781 HLCRDTCGSSSSSDNELAPPARAKSLPPSVTHSPLLHPRGFLRPSASLPEEAFASE 6840
Db 6781 HLCRDTCGSSSSSDNELAPPARAKSLPPSVTHSPLLHPRGFLRPSASLPEEAFASE 6840
Qy 6841 RSTAPAPPASPEGAGPPAAQGVPHSVIRSLFYHQAGSPHGMALPGSRHPPARRH 6900
Db 6841 RSTAPAPPASPEGAGPPAAQGVPHSVIRSLFYHQAGSPHGMALPGSRHPPARRH 6900
Qy 6901 LLKGGYTAGALPGLREPLMEHRVLEBAAREEATLLAKAPSPETALRLPASGTHLAPGH 6960
Db 6901 LLKGGYTAGALPGLREPLMEHRVLEBAAREEATLLAKAPSPETALRLPASGTHLAPGH 6960
Qy 6961 SHSLEHDSPTPRPSSACGAQLPSAPSGGAPIRDMGHPQSGKOLPSTGGHPGTAQPE 7020
Db 6961 SHSLEHDSPTPRPSSACGAQLPSAPSGGAPIRDMGHPQSGKOLPSTGGHPGTAQPE 7020
Qy 7021 RPSDPSWGOPAPFCHPKQSAQEGCSPPHAPVAPCPGSPFGSCKEAPLVSPSPFLGQ 7080
Db 7021 RPSDPSWGOPAPFCHPKQSAQEGCSPPHAPVAPCPGSPFGSCKEAPLVSPSPFLGQ 7080
Qy 7081 PQAPPAKASPPILDSXMGFCDISLPCRPKPGCSPGSSASQSSQVSLRVGSSQVGT 7140
Db 7081 PQAPPAKASPPILDSXMGFCDISLPCRPKPGCSPGSSASQSSQVSLRVGSSQVGT 7140
Qy 7141 EPGSLDAEGWTOEAEDLSSTPTLQRPQQAQWKEFSLGGRGGYAGVAGYGFAGGDA 7200
Db 7141 EPGSLDAEGWTOEAEDLSSTPTLQRPQQAQWKEFSLGGRGGYAGVAGYGFAGGDA 7200
Qy 7201 GGMGLQGPMMARIWAVSQSEEEQBEARAEQSEEQEAEPLPQVSARPVPEVGRA 7260
Db 7201 GGMGLQGPMMARIWAVSQSEEEQBEARAEQSEEQEAEPLPQVSARPVPEVGRA 7260
Qy 7261 PTRSPBPTMEDIGOVSLVQIRLSDGDAEAADTISLDISEVDPAVNLSDLDYDKLPP 7320
Db 7261 PTRSPBPTMEDIGOVSLVQIRLSDGDAEAADTISLDISEVDPAVNLSDLDYDKLPP 7320
Qy 7321 EFMIFRKVPKSAQPEPSPMAEELAEFEPTWPGELGHAGLEITESEEDVALLAE 7380
Db 7321 EFMIFRKVPKSAQPEPSPMAEELAEFEPTWPGELGHAGLEITESEEDVALLAE 7380
Qy 7381 AAVGKRKWSPSRSLFHPGRHLPLDEPAELGLRERVKASVEHISILKGRPEGLEKEG 7440
Db 7381 AAVGKRKWSPSRSLFHPGRHLPLDEPAELGLRERVKASVEHISILKGRPEGLEKEG 7440
Qy 7441 PPRKPGGLASFRSLGKSWDRAPTFLRELSDETIVVLOSVTLACQVSAQAQAATWSKDG 7500
Db 7441 PPRKPGGLASFRSLGKSWDRAPTFLRELSDETIVVLOSVTLACQVSAQAQAATWSKDG 7500
Qy 7501 APLESSRVLSATLKNFOLLITLVVAEDLGVTCSVSNALGTVTITGVLKKAERSSS 7560
Db 7501 APLESSRVLSATLKNFOLLITLVVAEDLGVTCSVSNALGTVTITGVLKKAERSSS 7560
Qy 7561 PCPDIGEVYADGVLLVWKPVESYGPVTIVQCSLEGGSWTTLASDIPDCCYLTSKLSRG 7620
Db 7561 PCPDIGEVYADGVLLVWKPVESYGPVTIVQCSLEGGSWTTLASDIPDCCYLTSKLSRG 7620
Qy 7621 TYTFTACVSKAGMGPYSPSPQVLLGGPSHLASEESQGSQAOLPSTKTFAQTQIQ 7680
Db 7621 TYTFTACVSKAGMGPYSPSPQVLLGGPSHLASEESQGSQAOLPSTKTFAQTQIQ 7680
Qy 7681 GRFSVVRQWEKASGRALAAKIIPIYHPKDKTAVLREYEALKGLRHPHLAQLHAYLSPRH 7740
Db 7681 GRFSVVRQWEKASGRALAAKIIPIYHPKDKTAVLREYEALKGLRHPHLAQLHAYLSPRH 7740
Qy 7741 LVLLLELCGSPBELLPCLABRASYSESEVKDYLWQMSATQYLNHGHILHLDLRSENMIIT 7800
Db 7741 LVLLLELCGSPBELLPCLABRASYSESEVKDYLWQMSATQYLNHGHILHLDLRSENMIIT 7800
Qy 7801 EYNLLKVDLGNASQISQEKVLPSPDKFDYLETMAPELLEGQCAVPEQTDIWAIGVTAFIM 7860
Db 7801 EYNLLKVDLGNASQISQEKVLPSPDKFDYLETMAPELLEGQCAVPEQTDIWAIGVTAFIM 7860

RESULT 2

US-10-080-334-290
; Sequence 290, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkets, Richard A
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zernhusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/276,400
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810

; PRIOR FILING DATE: 2001-02-23
 ; NUMBER OF SEQ ID NOS: 388
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 290
 ; LENGTH: 6620
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-080-334-290

Query Match 77.9%; Score 32134; DB 15; Length 6620;
 Best Local Similarity 98.1%; Pred. No. 0;
 Matches 6250; Conservative 15; Mismatches 57; Indels 48; Gaps 9;

| | | | |
|----|-----|--|-----|
| QY | 1 | MDQPFSGAPFLTRPKAFVSVGKDATLSQIVGNPTPOVSWKDDQOPVTAGARFLAQ | 60 |
| DB | 1 | MDQPFSGAPFLTRPKAFVSVGKDATLSQIVGNPTPOVSWKDDQOPVTAGARFLAQ | 60 |
| QY | 61 | DGDLVRLTILDLALGDSQYVCRARNATGEAFAGVLOYDAEACAOAHPFLRPTSIIR | 120 |
| DB | 61 | DGDLVRLTILDLALGDSQYVCRARNATGEAFAGVLOYDAEACAOAHPFLRPTSIIR | 120 |
| QY | 121 | VREGSEATFCRVGSGPRPAVSWKDGRRLLGEPDGPVRVVEELGEASALIRIARPRDGG | 180 |
| DB | 121 | VREGSEATFCRVGSGPRPAVSWKDGRRLLGEPDGPVRVVEELGEASALIRIARPRDGG | 180 |
| QY | 181 | TYEVRAENPLCAASAAALVDSADATASPGTSTAALLAHQORREANRAEGAPASPP | 240 |
| DB | 181 | TYEVRAENPLCAASAAALVDSADATASPGTSTAALLAHQORREANRAEGAPASPP | 240 |
| QY | 241 | STGTRTCTVTGKARLSQYVTGPKPETVWKDGQLVTEGRRHVVEDAQENFVLKILF | 300 |
| DB | 241 | STGTRTCTVTGKARLSQYVTGPKPETVWKDGQLVTEGRRHVVEDAQENFVLKILF | 300 |
| QY | 301 | CKQSDRGLYCTASNLVQGTYSVVLVVRBPAPVFKKRLQDLEVRKESATFLCEVPQPS | 360 |
| DB | 301 | CKQSDRGLYCTASNLVQGTYSVVLVVRBPAPVFKKRLQDLEVRKESATFLCEVPQPS | 360 |
| QY | 361 | TEAAWKEETFLWASAKYIEEGTERLTVRNVSADDDAVYICETPEGSRTVAELAVQG | 420 |
| DB | 361 | TEAAWKEETFLWASAKYIEEGTERLTVRNVSADDDAVYICETPEGSRTVAELAVQG | 420 |
| QY | 421 | NLLRKLPRKTAVRVGD7AMFCVELAVPVPVHWRNQEEVVAAGRVVAISAEGRTHLTIS | 480 |
| DB | 421 | NLLRKLPRKTAVRVGD7AMFCVELAVPVPVHWRNQEEVVAAGRVVAISAEGRTHLTIS | 480 |
| QY | 481 | QCCLLEDVGOAFMAGDCQSTRFCVSAAPRPPLOPPVDVVKARMESSTVLSWSPPHGE | 540 |
| DB | 481 | QCCLLEDVGOAFMAGDCQSTRFCVSAAPRPPLOPPVDVVKARMESSTVLSWSPPHGE | 540 |
| QY | 541 | RPVTIDGLYVKKKLGTYTIRCHEAEWATPELTVDVAEENGFQFRYSALNSFCQSPY | 600 |
| DB | 541 | RPVTIDGLYVKKKLGTYTIRCHEAEWATPELTVDVAEENGFQFRYSALNSFCQSPY | 600 |
| QY | 601 | LEPGTTHLAPKLAVRTPLKAVQAVEGGEVTFSDTLTVASAGFWFLDQALKASSVYEIH | 660 |
| DB | 601 | LEPGTTHLAPKLAVRTPLKAVQAVEGGEVTFSDTLTVASAGFWFLDQALKASSVYEIH | 660 |
| QY | 661 | CDTRTHLTITREVPASLHGQALKFVANGIBESIRMEVRAAPGLTANKPPAAAREVLARL | 720 |
| DB | 661 | CDTRTHLTITREVPASLHGQALKFVANGIBESIRMEVRAAPGLTANKPPAAAREVLARL | 720 |
| QY | 721 | HEEAQLLAELSDQAAAVTNLKDQRTLSPGPKYEVQASAGRRVLLVRDVAARDAGLYECVS | 780 |
| DB | 721 | HEEAQLLAELSDQAAAVTNLKDQRTLSPGPKYEVQASAGRRVLLVRDVAARDAGLYECVS | 780 |
| QY | 781 | RGRIAYQLSVQGLARFLHKDMAGSCVDVAGGPAQFECETSEAHVHVHWYKDGVELGHS | 840 |
| DB | 781 | RGRIAYQLSVQGLARFLHKDMAGSCVDVAGGPAQFECETSEAHVHVHWYKDGVELGHS | 840 |
| QY | 841 | GERFLOEDVGTTRHLVAATVTRDEGTYSCRVGEDSVDFRLRVSPKPVFAKEQLARRKL | 900 |
| DB | 841 | GERFLOEDVGTTRHLVAATVTRDEGTYSCRVGEDSVDFRLRVSPKPVFAKEQLARRKL | 900 |

| | | | |
|----|------|--|------|
| QY | 901 | QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMTEATCTRRLLVVOQAGQADAGEY | 960 |
| DB | 901 | QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMTEATCTRRLLVVOQAGQADAGEY | 960 |
| QY | 961 | SCEAGGQRLSFHLDVKEPKVFAKQVAHSEVQAEAGANATLSCEVAQAQAEVWYKDGK | 1020 |
| DB | 961 | SCEAGGQRLSFHLDVKEPKVFAKQVAHSEVQAEAGANATLSCEVAQAQAEVWYKDGK | 1020 |
| QY | 1021 | KLSSSLKHVVEAKGCRRLVVOQAQKTDAGDYSCEARGQVSRFLHITPEKMFVFAKESV | 1080 |
| DB | 1021 | KLSSSLKHVVEAKGCRRLVVOQAQKTDAGDYSCEARGQVSRFLHITPEKMFVFAKESV | 1080 |
| QY | 1081 | HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSSSSKVMEVKGCTRRLLVPOAGKAD | 1140 |
| DB | 1081 | HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSSSSKVMEVKGCTRRLLVPOAGKAD | 1140 |
| QY | 1141 | AGEYSCEAGGQVSRFLHITPEKGVFAKESQVHNEVQAEAGTTAMLSCEVAQOPQTEVTWY | 1200 |
| DB | 1141 | AGEYSCEAGGQVSRFLHITPEKGVFAKESQVHNEVQAEAGTTAMLSCEVAQOPQTEVTWY | 1200 |
| QY | 1201 | KDGKLSSSSKVMEVKGCTRRLLVVOQVKGADAGEYSCEAGGQVSRFLHITPEKAVFAK | 1260 |
| DB | 1201 | KDGKLSSSSKVMEVKGCTRRLLVVOQVKGADAGEYSCEAGGQVSRFLHITPEKAVFAK | 1260 |
| QY | 1261 | EOLVHNEVTRTEAGSATLSCEVAQAQTEVTWYKDGKLSSSSKVRIEAGCMRQLVVOQA | 1320 |
| DB | 1261 | EOLVHNEVTRTEAGSATLSCEVAQAQTEVTWYKDGKLSSSSKVRIEAGCMRQLVVOQA | 1320 |
| QY | 1321 | GOADAGEVTCRAGGQRLSFHLDVSEPKAVFAKEOALHAKVQAEAGATATLSCEVAQAQTE | 1380 |
| DB | 1321 | GOADAGEVTCRAGGQRLSFHLDVSEPKAVFAKEOALHAKVQAEAGATATLSCEVAQAQTE | 1380 |
| QY | 1381 | VTWYKDGKLSSSSKVMEVKGCTRRLLVVOQAQADTGEYSCEAGGQRLSFSLDVAPKV | 1440 |
| DB | 1381 | VTWYKDGKLSSSSKVMEVKGCTRRLLVVOQAQADTGEYSCEAGGQRLSFSLDVAPKV | 1440 |
| QY | 1441 | VFAKEQPVHREVQAGASTTSLCEVAQAQTEVTWYKDGKLSFSSKVRMEAVGCTRRLLV | 1500 |
| DB | 1441 | VFAKEQPVHREVQAGASTTSLCEVAQAQTEVTWYKDGKLSFSSKVRMEAVGCTRRLLV | 1500 |
| QY | 1501 | VQAGQADAGEYSCEAGSQRSLFHLVAPKAVFAKQAPASREVQAEAGTATLSCEVAQ | 1560 |
| DB | 1501 | VQAGQADAGEYSCEAGSQRSLFHLVAPKAVFAKQAPASREVQAEAGTATLSCEVAQ | 1560 |
| QY | 1561 | AQTEVTWYKDGKLSSSSKVMEVKGCTRRLLVVOQAQADAGEYSCKAGQRLSFHLV | 1620 |
| DB | 1561 | AQTEVTWYKDGKLSSSSKVMEVKGCTRRLLVVOQAQADAGEYSCKAGQRLSFHLV | 1620 |
| QY | 1621 | BPKVFAKEQPAHREVQAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRVEAVGCT | 1680 |
| DB | 1621 | BPKVFAKEQPAHREVQAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRVEAVGCT | 1680 |
| QY | 1681 | RRLVVOQAQADAGEYSCEAGGQRLSFHLVAPLEPOISERPCCRREPLVVKHEHDIILTA | 1740 |
| DB | 1681 | RRLVVOQAQADAGEYSCEAGGQRLSFHLVAPLEPOISERPCCRREPLVVKHEHDIILTA | 1740 |
| QY | 1741 | TLATPSAANTVWLKDGVEIRSRKHETASQGDTHLTTHGAQVLDLSAISCYVGAEGQDF | 1800 |
| DB | 1741 | TLATPSAANTVWLKDGVEIRSRKHETASQGDTHLTTHGAQVLDLSAISCYVGAEGQDF | 1800 |
| QY | 1801 | PVQVEEVAAKFCRLLEPVCGLGTTVTLACELSPACAEVWVRGNTQPRVGKQFQVMAEG | 1860 |
| DB | 1801 | PVQVEEVAAKFCRLLEPVCGLGTTVTLACELSPACAEVWVRGNTQPRVGKQFQVMAEG | 1860 |
| QY | 1861 | PVRSLTIVGLRAEDAGEYVCESDHHTSAQLTIVSVPRVVKFMSGLSTVVAEAGEATFQC | 1920 |
| DB | 1861 | PVRSLTIVGLRAEDAGEYVCESDHHTSAQLTIVSVPRVVKFMSGLSTVVAEAGEATFQC | 1920 |
| QY | 1921 | VWSPDVAVWVRDQALLQPSKFAISQSGASHLTIISDLVLEDAQOITVVEAEGSSAA | 1980 |
| DB | 1921 | VWSPDVAVWVRDQALLQPSKFAISQSGASHLTIISDLVLEDAQOITVVEAEGSSAA | 1980 |
| QY | 1981 | LRVREAPVLFKKLEBPQTVREERSVTLEVLTRFWPELRTNRNATALAPKGNVIEHAEGA | 2040 |

| | | | | | | | | | | | |
|------|----|--------|------------|------------|--------------------|---------------------|--------------------------|---------------------|--------------|------------------|----------------|
| 1981 | Db | LVR | REAPVL | FKKKLEPOTV | SEERSSVTLEBELTFPWP | ELRWTRNATALAPCKNVEI | HAEGA | 2040 | | | |
| 2041 | Qy | RHRLV | HNVGADRG | FFGCGT | PDKQAKLT | VMROVRLVRGLQ | VAEAREOGTATMEVOL | 2100 | | | |
| 2041 | Db | RHRLV | HNVGADRG | FFGCGT | PDKQAKLT | VMROVRLVRGLQ | VAEAREOGTATMEVOL | 2100 | | | |
| 2101 | Qy | SHADV | GSWTRDGLR | FQOGPT | CHLAVRGP | MHTLTLSGLR | PEDSGLMVFKAEGVHTSARLV | 2160 | | | |
| 2101 | Db | SHADV | GSWTRDGLR | FQOGPT | CHLAVRGP | MHTLTLSGLR | PEDSGLMVFKAEGVHTSARLV | 2160 | | | |
| 2161 | Qy | VTLPV | SFSPLODV | VTTEKEKVT | LBCELSRP | NVDVRWLKDGVEL | RAGKTMATAAOGACR | 2220 | | | |
| 2161 | Db | VTLPV | SFSPLODV | VTTEKEKVT | LBCELSRP | NVDVRWLKDGVEL | RAGKTMATAAOGACR | 2220 | | | |
| 2221 | Qy | SLTIYR | CEFADQGVYV | CDADHA | QSSASVKVQGR | TYTLIYRVL | AEADAGEIQFVAE | NAESR 2280 | | | |
| 2221 | Db | SLTIYR | CEFADQGVYV | CDADHA | QSSASVKVQGR | TYTLIYRVL | AEADAGEIQFVAE | NAESR 2280 | | | |
| 2281 | Qy | AQLRV | KELPVTILVR | PLRDKTAME | KHRGVL | ECQVSR | ASAOVRWFKGSOELQPGKYELVSD | 2340 | | | |
| 2281 | Db | AQLRV | KELPVTILVR | PLRDKTAME | KHRGVL | ECQVSR | ASAOVRWFKGSOELQPGKYELVSD | 2340 | | | |
| 2341 | Qy | GLYRK | LIISDV | HAEDETYTC | DAGDVKTSAQ | FFVEQSIITVR | GLQDVTVM | PAPAWFEC 2400 | | | |
| 2341 | Db | GLYRK | LIISDV | HAEDETYTC | DAGDVKTSAQ | FFVEQSIITVR | GLQDVTVM | PAPAWFEC 2400 | | | |
| 2401 | Qy | ETSI | PSVRPKW | LIGKTVLQ | AGNVGL | QEBQGTVHRL | MLRFTCSMTGPHVFTV | GKSRSSA 2460 | | | |
| 2401 | Db | ETSI | PSVRPKW | LIGKTVLQ | AGNVGL | QEBQGTVHRL | MLRFTCSMTGPHVFTV | GKSRSSA 2460 | | | |
| 2461 | Qy | RLVSDI | PVVLTR | PLEPKT | REQLQS | VVLS | CDFRPAPKAVQWYKDDT | PLSPSEKFKMSLEG 2520 | | | |
| 2461 | Db | RLVSDI | PVVLTR | PLEPKT | REQLQS | VVLS | CDFRPAPKAVQWYKDDT | PLSPSEKFKMSLEG 2520 | | | |
| 2521 | Qy | QMAEL | RILRLMP | DAGVYR | COAGSAH | SSTETVVEAR | EVTVTGPLQDAEATEG | WASFSC 2580 | | | |
| 2521 | Db | QMAEL | RILRLMP | DAGVYR | COAGSAH | SSTETVVEAR | EVTVTGPLQDAEATEG | WASFSC 2580 | | | |
| 2581 | Qy | LSHED | EEVWSL | NGMPLYNDS | PFHEI | SHKGR | HTLVLSIORADAGIVR | ASSLKVSARSARL 2640 | | | |
| 2581 | Db | LSHED | EEVWSL | NGMPLYNDS | PFHEI | SHKGR | HTLVLSIORADAGIVR | ASSLKVSARSARL 2640 | | | |
| 2641 | Qy | EVVRK | PVFLKAL | DDL | LSAEB | EGTLALQCE | YSDPEAHVVRKDG | VQLGSPDKYD | FLHTAGT 2700 | | |
| 2641 | Db | EVVRK | PVFLKAL | DDL | LSAEB | EGTLALQCE | YSDPEAHVVRKDG | VQLGSPDKYD | FLHTAGT 2700 | | |
| 2701 | Qy | RGLV | VHDVSP | DAGLYT | CHVGS | ETRARVRV | DHLHGCIITKRLT | MEVLEGE | SCSFEVL 2760 | | |
| 2701 | Db | RGLV | VHDVSP | DAGLYT | CHVGS | ETRARVRV | DHLHGCIITKRLT | MEVLEGE | SCSFEVL 2760 | | |
| 2761 | Qy | HESAS | DPMVTVG | KTVG | SGSSSRFOAT | QGRKYILV | VEAAPSDAGEV | VVFSVRGLTSKASL 2820 | | | |
| 2761 | Db | HESAS | DPMVTVG | KTVG | SGSSSRFOAT | QGRKYILV | VEAAPSDAGEV | VVFSVRGLTSKASL 2820 | | | |
| 2821 | Qy | IVRER | PAALIKPL | BDQWAP | GEDVEL | LBCEL | SRACTPVH | WKORKAIRKSOKY | DVVCBGT 2880 | | |
| 2821 | Db | IVRER | PAALIKPL | BDQWAP | GEDVEL | LBCEL | SRACTPVH | WKORKAIRKSOKY | DVVCBGT 2880 | | |
| 2881 | Qy | AMLVI | RASLKD | AGEYTC | VEASKSTASL | HVEEK | XANCFTEBLTNL | VEEKGTA | VFTCKTE 2940 | | |
| 2881 | Db | AMLVI | RASLKD | AGEYTC | VEASKSTASL | HVEEK | XANCFTEBLTNL | VEEKGTA | VFTCKTE 2940 | | |
| 2941 | Qy | HPAA | TVTW | RKGLLE | LRASG | KHPQ | SOEGLTURLT | LTISALEK | ADSDTYTCDI | GOAQSRAOLLY 3000 | |
| 2941 | Db | HPAA | TVTW | RKGLLE | LRASG | KHPQ | SOEGLTURLT | LTISALEK | ADSDTYTCDI | GOAQSRAOLLY 3000 | |
| 3001 | Qy | QGRV | HHII | EDLED | VNDV | QEGSSAT | FRCLIS | PANYBPV | HWFLDKT | PLHANELNEID | DAQPGY 3060 |
| 3001 | Db | QGRV | HHII | EDLED | VNDV | QEGSSAT | FRCLIS | PANYBPV | HWFLDKT | PLHANELNEID | DAQPGY 3060 |
| 3061 | Qy | HVLTR | QLAL | KDSGT | TYFEA | GDQ | BSAALRYTER | KPSVFSR | BLTDATTE | GED | BLTVLCETS 3120 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|------|----|------|-------|----|-----|-----|-----|----|----|----|----|----|----|----|----|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|------|------|---|------|---|---|---|------|
| 3061 | Db | HVLT | RLQAL | KD | SGT | TYF | EAG | DQ | RA | SA | AL | RV | TE | KP | SV | F | RE | L | T | D | A | T | I | T | E | G | E | D | L | T | V | C | E | T | S | 3120 | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3121 | Qy | TC | D | P | M | C | W | T | K | D | G | K | T | L | R | G | S | A | R | C | Q | L | S | H | E | G | H | R | A | Q | L | L | I | T | G | A | T | Q | D | S | G | R | Y | K | E | A | G | A | C | S | S | I | V | R | 3180 | | | | | | |
| 3121 | Db | TC | D | P | M | C | W | T | K | D | G | K | T | L | R | G | S | A | R | C | Q | L | S | H | E | G | H | R | A | Q | L | L | I | T | G | A | T | Q | D | S | G | R | Y | K | E | A | G | A | C | S | S | I | V | R | 3180 | | | | | | |
| 3181 | Qy | H | A | P | V | R | F | O | E | A | L | K | O | L | E | V | L | E | G | A | A | T | L | R | C | V | L | S | S | V | A | P | V | K | C | Y | G | N | N | V | L | R | P | G | D | K | Y | S | L | R | O | E | G | A | M | L | 3240 | | | | |
| 3181 | Db | H | A | P | V | R | F | O | E | A | L | K | O | L | E | V | L | E | G | A | A | T | L | R | C | V | L | S | S | V | A | P | V | K | C | Y | G | N | N | V | L | R | P | G | D | K | Y | S | L | R | O | E | G | A | M | L | 3240 | | | | |
| 3241 | Qy | L | V | E | N | L | R | P | D | S | G | R | Y | S | C | S | F | G | D | T | T | S | A | T | L | T | V | T | A | L | P | A | Q | F | I | G | K | L | R | N | K | E | A | T | E | G | A | T | L | R | C | E | L | S | K | T | 3300 | | | | |
| 3241 | Db | L | V | E | N | L | R | P | D | S | G | R | Y | S | C | S | F | G | D | T | T | S | A | T | L | T | V | T | A | L | P | A | Q | F | I | G | K | L | R | N | K | E | A | T | E | G | A | T | L | R | C | E | L | S | K | T | 3300 | | | | |
| 3301 | Qy | A | P | V | E | W | R | K | G | S | T | L | R | D | R | Y | C | L | R | O | G | A | M | C | E | L | O | I | R | G | L | A | M | D | A | A | E | Y | S | C | V | C | G | E | R | T | S | A | L | T | I | R | M | 3360 | | | | | | | |
| 3301 | Db | A | P | V | E | W | R | K | G | S | T | L | R | D | R | Y | C | L | R | O | G | A | M | C | E | L | O | I | R | G | L | A | M | D | A | A | E | Y | S | C | V | C | G | E | R | T | S | A | L | T | I | R | M | 3360 | | | | | | | |
| 3361 | Qy | P | A | H | E | I | G | L | R | H | O | E | S | I | E | G | A | T | L | R | C | E | L | S | K | A | A | P | V | E | W | R | K | G | R | E | S | L | R | D | G | R | H | S | L | R | O | D | G | A | V | C | E | L | Q | I | 3420 | | | | |
| 3361 | Db | P | A | H | E | I | G | L | R | H | O | E | S | I | E | G | A | T | L | R | C | E | L | S | K | A | A | P | V | E | W | R | K | G | R | E | S | L | R | D | G | R | H | S | L | R | O | D | G | A | V | C | E | L | Q | I | 3420 | | | | |
| 3421 | Qy | G | L | A | V | A | D | A | G | E | Y | S | C | V | C | G | E | R | T | S | A | T | L | T | V | X | A | L | P | A | K | F | T | E | G | L | R | N | E | E | A | V | E | G | A | T | A | M | L | W | C | E | L | S | K | V | A | P | V | E | 3480 |
| 3421 | Db | G | L | A | V | A | D | A | G | E | Y | S | C | V | C | G | E | R | T | S | A | T | L | T | V | X | A | L | P | A | K | F | T | E | G | L | R | N | E | E | A | V | E | G | A | T | A | M | L | W | C | E | L | S | K | V | A | P | V | E | 3480 |
| 3481 | Qy | W | R | K | P | E | N | L | R | D | G | R | Y | L | R | O | E | G | T | R | C | E | L | O | I | C | G | L | A | M | A | D | A | G | E | Y | L | C | V | | | | | | | | | | | | | | | | | | | | | | |

QY 4201 SNEVTEAVRDRGRIHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVITLPLEQDVQ 4260
 DB SNEVTEAVRDRGRIHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVITLPLEQDVQ 4260
 QY 4261 LSEGQDASFQCRLSRASQCEARWALGGVPLQANEMNDITVBOGTLHLTLHKVTELDAGT 4320
 DB LSEGQDASFQCRLSRASQCEARWALGGVPLQANEMNDITVBOGTLHLTLHKVTELDAGT 4320
 QY 4321 VSFHVGTCSSAQLKVTAKNTVWGLENVVALEGEALFEQOLSQPEVAATHWLLDDEPV 4380
 DB VSFHVGTCSSAQLKVTAKNTVWGLENVVALEGEALFEQOLSQPEVAATHWLLDDEPV 4380
 QY 4381 RTSNAEVVFFENGLRHLLKLNLRPQDSCRVTFLAGDMVTSAFITVRGWELEILEPLKN 4440
 DB RTSNAEVVFFENGLRHLLKLNLRPQDSCRVTFLAGDMVTSAFITVRGWELEILEPLKN 4440
 QY 4441 AAVRAGAQAARTCTILSEAVPVGEASWYINGAAVOPDDSDMTVTADGSHQALLRSQAQPHH 4500
 DB AAVRAGAQAARTCTILSEAVPVGEASWYINGAAVOPDDSDMTVTADGSHQALLRSQAQPHH 4500
 QY 4501 AGEVTFACRDVAVASARLTVLGLPDPPEDAEVVAHSSHTVLSWAAPMSDGGGLCGYRVE 4560
 DB AGEVTFACRDVAVASARLTVLGLPDPPEDAEVVAHSSHTVLSWAAPMSDGGGLCGYRVE 4560
 QY 4561 VKEGATGQWRLCHELVPGPECVVDGLAPGETYRFRVAAGVGVGAGEPVHLPQTVRLAEP 4620
 DB VKEGATGQWRLCHELVPGPECVVDGLAPGETYRFRVAAGVGVGAGEPVHLPQTVRLAEP 4620
 QY 4621 KPVPPOPSAPRSROVAGEDYSLELVVAZAGEVIMHKGMRLOPGGRFEVSGRQOML 4680
 DB KPVPPOPSAPRSROVAGEDYSLELVVAZAGEVIMHKGMRLOPGGRFEVSGRQOML 4680
 QY 4681 VIKGFTABDQGEVHCGLAQGSICPAATFQVALSPASVDEAPQSLPPEAAQEGDLHLW 4740
 DB VIKGFTABDQGEVHCGLAQGSICPAATFQVALSPASVDEAPQSLPPEAAQEGDLHLW 4740
 QY 4741 EALARKRMSREPTLDSISELPEEDGRSQRLOPQAEVAVDLSGYSYTADELARTGDAL 4800
 DB EALARKRMSREPTLDSISELPEEDGRSQRLOPQAEVAVDLSGYSYTADELARTGDAL 4800
 QY 4801 SHTSSDDESRAGTSLVTLKACRPGTSPILASKVGAAPAPSVKPOQOQEPPLAARPPPLG 4860
 DB SHTSSDDESRAGTSLVTLKACRPGTSPILASKVGAAPAPSVKPOQOQEPPLAARPPPLG 4860
 QY 4861 DLSTKDLDGDPMDXAAVKIQAAFKGYKVRKEMKQOQEGMFSTFGDTEAVQGDALRLCV 4920
 DB DLSTKDLDGDPMDXAAVKIQAAFKGYKVRKEMKQOQEGMFSTFGDTEAVQGDALRLCV 4920
 QY 4921 VASADVAVRVLKDGVELTDGRHHIIPOLGDGTCSLLIAGLDRADAGCYTCQVSNKFGOV 4980
 DB VASADVAVRVLKDGVELTDGRHHIIPOLGDGTCSLLIAGLDRADAGCYTCQVSNKFGOV 4980
 QY 4981 THSAACVVVSGSESAESSSGEELDDAFRAARLRLHRLFRTKSPAESVDEELFLSADGGA 5040
 DB THSAACVVVSGSESAESSSGEELDDAFRAARLRLHRLFRTKSPAESVDEELFLSADGGA 5040
 QY 5041 EPEEPADQWYREDEHFCIRFEALTEARQAVTRFOEMPATLIGVIEIKLVEQGRVEM 5100
 DB EPEEPADQWYREDEHFCIRFEALTEARQAVTRFOEMPATLIGVIEIKLVEQGRVEM 5100
 QY 5101 CISKETPAVVPVPPPLSLTSDAAPVFLTELQNEVQDGPVPSFCVVTGQMPSPVRWF 5160
 DB CISKETPAVVPVPPPLSLTSDAAPVFLTELQNEVQDGPVPSFCVVTGQMPSPVRWF 5160
 QY 5161 KDGKLLSEDDHYMINEQCGHQHIIITAVVPADWGVYRCLAENSMGVSTKAEIRVDLTS 5220
 DB KDGKLLSEDDHYMINEQCGHQHIIITAVVPADWGVYRCLAENSMGVSTKAEIRVDLTS 5220
 QY 5221 TDYDTAADATESSSYFSAQGYLSSREOGETSTTDEQQLPQVVEELRDQLVAFGTBLAKP 5280
 DB TDYDTAADATESSSYFSAQGYLSSREOGETSTTDEQQLPQVVEELRDQLVAFGTBLAKP 5280

QY 5281 QLKVKGYGYPAPRLYWFKDGOPLTASAHIRMTGKKILHTLEIISVTRSDSGQAAVYISNMG 5340
 DB QLKVKGYGYPAPRLYWFKDGOPLTASAHIRMTGKKILHTLEIISVTRSDSGQAAVYISNMG 5340
 QY 5341 AAYSSARLLVRGPDPEEKPPASDVHEQLVPPRMLERFTPKVKYKGGSSITFSVKVEGRVVP 5400
 DB AAYSSARLLVRGPDPEEKPPASDVHEQLVPPRMLERFTPKVKYKGGSSITFSVKVEGRVVP 5400
 QY 5401 TVHMLREERBERGVLMIGDTPGTYTVASSAQOHSVLVLDVGRHQGTTCIASNAAGQALC 5460
 DB TVHMLREERBERGVLMIGDTPGTYTVASSAQOHSVLVLDVGRHQGTTCIASNAAGQALC 5460
 QY 5461 SASLHVSGLPKVBEQEKVKEALISTFLOTTQOASAOGLETASPADLGGQKEEPLAAKE 5520
 DB SASLHVSGLPKVBEQEKVKEALISTFLOTTQOASAOGLETASPADLGGQKEEPLAAKE 5520
 QY 5521 ALGHLSLAEVGTETPLQKLTQOITEMVSAKITQAKLQVPGGSDSDSKTSPASPRHGSR 5580
 DB ALGHLSLAEVGTETPLQKLTQOITEMVSAKITQAKLQVPGGSDSDSKTSPASPRHGSR 5580
 QY 5581 PSSSQESSSESDGDARGEIYVVTADYLPQAEQDAITLREGQVVEVLDAHPPLRW 5640
 DB PSSSQESSSESDGDARGEIYVVTADYLPQAEQDAITLREGQVVEVLDAHPPLRW 5640
 QY 5641 LVRTKPTKSSPSRQWVSPAYLDRRLKLSPEWGAAPAEPEFPGEAVERSEYKARLSSVTOE 5700
 DB LVRTKPTKSSPSRQWVSPAYLDRRLKLSPEWGAAPAEPEFPGEAVERSEYKARLSSVTOE 5700
 QY 5701 LSSSQAFVEELQFLQSHLQHLERCCHVPITAVAGOKAVIPRNVRDTCGRHSSFLQELQ 5760
 DB LSSSQAFVEELQFLQSHLQHLERCCHVPITAVAGOKAVIPRNVRDTCGRHSSFLQELQ 5760
 QY 5761 CDTDDVAMCFIKNOAFAFEQYLEFLVGRVQAESVWVSTAIQEFYKKAEEALLAGDSQP 5820
 DB CDTDDVAMCFIKNOAFAFEQYLEFLVGRVQAESVWVSTAIQEFYKKAEEALLAGDSQP 5820
 QY 5821 PPPLOHLYEQPVERVQYQALLKELIRNKARNQNCALLQOAVVVSALPQRAENKLHV 5880
 DB PPPLOHLYEQPVERVQYQALLKELIRNKARNQNCALLQOAVVVSALPQRAENKLHV 5880
 QY 5881 SLMENYPTGLAALGPRIQGHFIVMEGAPGARMKPMKGNHRVFLFRNHLVTCRPRDSRT 5940
 DB SLMENYPTGLAALGPRIQGHFIVMEGAPGARMKPMKGNHRVFLFRNHLVTCRPRDSRT 5940
 QY 5941 DTVSIVFRNMKLSIDLNDQVGGDDRAFEVWQEREDSVRYKLLQARTAIKSSWVKEIC 6000
 DB DTVSIVFRNMKLSIDLNDQVGGDDRAFEVWQEREDSVRYKLLQARTAIKSSWVKEIC 6000
 QY 6001 GICQRLALPVNRPPDFEELADCTAELGETVKLACRVGTGPKPVI SWYKDGKAVQVDPHH 6060
 DB GICQRLALPVNRPPDFEELADCTAELGETVKLACRVGTGPKPVI SWYKDGKAVQVDPHH 6060
 QY 6061 ILIEDPDGSCALILDSLTGVDGQVCMFAASAAGNCSTGLKILVQVPPRFVNVKVRASFFV 6120
 DB ILIEDPDGSCALILDSLTGVDGQVCMFAASAAGNCSTGLKILVQVPPRFVNVKVRASFFV 6120
 QY 6121 EGEDAQFTCTTEGAPYPOIRWYKDGALLTTGNKFQTLSEPRSGLLVIVRAASKEDLGLY 6180
 DB EGEDAQFTCTTEGAPYPOIRWYKDGALLTTGNKFQTLSEPRSGLLVIVRAASKEDLGLY 6180
 QY 6181 ECELVNRLGSAARAEILRIQSPMLQAOEQCHREQLVAAVEVTEQETKVPKKTVIIETIT 6240
 DB ECELVNRLGSAARAEILRIQSPMLQAOEQCHREQLVAAVEVTEQETKVPKKTVIIETIT 6240
 QY 6241 TVVKSPPRCQR - SPSKSPRSRSPRSCSASPLRPGLLAPDLLYLPQAGQPRPEABPGQKPV 6299
 DB TVVKSPPRCQR - SPSKSPRSRSPRSCSASPLRPGLLAPDLLYLPQAGQPRPEABPGQKPV 6299
 QY 6299 - PVTGTSEAPVPRV - - - - - POPLLHEGPE - - - - - QEPFAIARAQEWTPIRMGA 6317
 DB VPTLVVTEAEAHSPALPGLSGQPKVVEVEETIEVRVKWKGPGQVSPTE - - - - - VPRSSSGH 6357
 QY 6357 AW - - - - - PGAGTG 6325


```

Db      6358 LFTLPATPG 6367
;      |||
;
RESULT 3
US-10-408-765A-2291
; Sequence 2291, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Chosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2291:
; LENGTH: 6620
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2291

Query Match      77.9%; Score 32134; DB 16; Length 6620;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 6250; Conservative 15; Mismatches 57; Indels 48; Gaps 9;

Qy      1 MDQPFSGAPFLTRPKAFVVSVKGDATLSQIVGNTPPOVSWEKDDQPPVTAGARFLAQ 60
Db      1 MDQPFSGAPFLTRPKAFVVSVKGDATLSQIVGNTPPOVSWEKDDQPPVTAGARFLAQ 60

Qy      61 DGDLYRLTLDLALGDSQYVCARNAGIIFAFAVGLQVDAEACACQAPHFLLRPTISR 120
Db      61 DGDLYRLTLDLALGDSQYVCARNAGIIFAFAVGLQVDAEACACQAPHFLLRPTISR 120

Qy      121 VREGSEATFRGCGSPAPVSWKGRRLCEPPGPRVVRVEELGEASALRIARAPRDGG 180
Db      121 VREGSEATFRGCGSPAPVSWKGRRLCEPPGPRVVRVEELGEASALRIARAPRDGG 180

Qy      181 TYEVRANPLGAASAAAALVWDSADATASPGTSTAALLAHLQRRREAVRAEGAPASPP 240
Db      181 TYEVRANPLGAASAAAALVWDSADATASPGTSTAALLAHLQRRREAVRAEGAPASPP 240

Qy      241 STGTRTCTVTGKHARLSCYVTGPKPETVWKDQQLVTEGRRHVWYEDAQENFVKILF 300
Db      241 STGTRTCTVTGKHARLSCYVTGPKPETVWKDQQLVTEGRRHVWYEDAQENFVKILF 300

Qy      301 CKQSDRGLYTCTASNLVGQTYSSVLVVRREPAVPFKRLQDLEVREKESATFLCEVPQPS 360
Db      301 CKQSDRGLYTCTASNLVGQTYSSVLVVRREPAVPFKRLQDLEVREKESATFLCEVPQPS 360

Qy      361 TEAAMFKETRLWASAKYIGIEEGTERLTVRNVSADDDAVYICETPEGSRVTAELAVQG 420
Db      361 TEAAMFKETRLWASAKYIGIEEGTERLTVRNVSADDDAVYICETPEGSRVTAELAVQG 420

Qy      421 NLLRLPKRTAVRUGDTAMFCVELAVPVGPHWLRNQEVEVAGRVVAISAGTRHTLTIS 480
Db      421 NLLRLPKRTAVRUGDTAMFCVELAVPVGPHWLRNQEVEVAGRVVAISAGTRHTLTIS 480

Qy      481 QCCLDEVQVAFMAGDCQSTSTRFCVSAAPRKPLOPPVDPVVKARMESSVILSWSPPHGE 540
Db      481 QCCLDEVQVAFMAGDCQSTSTRFCVSAAPRKPLOPPVDPVVKARMESSVILSWSPPHGE 540

Qy      541 RPVTIDGVLVEKKLGTWIRCHEAEWVATPELTADVAEENQFQFRVSAINSGOSP 600
Db      541 RPVTIDGVLVEKKLGTWIRCHEAEWVATPELTADVAEENQFQFRVSAINSGOSP 600

601 LEFFPGTVHLAPKLAVRTPLKAVQAVEGGEVTFSDLTVASAGWFLDGOALKASSVYETH 660
601 LEFFPGTVHLAPKLAVRTPLKAVQAVEGGEVTFSDLTVASAGWFLDGOALKASSVYETH 660
661 CDTRHTLTIREVPASLHGAQLKFVANGIESSIRVEVRAAPGLTANKPPAAAAREVLAARL 720
661 CDTRHTLTIREVPASLHGAQLKFVANGIESSIRVEVRAAPGLTANKPPAAAAREVLAARL 720
721 HEEAQLLAELSQAQAATVWLKDGRTLSGPKPKYEVQASAGRRVLLVRDVARDDAGLYECVS 780
721 HEEAQLLAELSQAQAATVWLKDGRTLSGPKPKYEVQASAGRRVLLVRDVARDDAGLYECVS 780
781 RGRRIAYQLSVQGLARFLHDKMAGSCVDVAGGPAQFECETSEAHVHVHWYKXGVELGHS 840
781 RGRRIAYQLSVQGLARFLHDKMAGSCVDVAGGPAQFECETSEAHVHVHWYKXGVELGHS 840
841 GERFLOEDVGTGTHRLVAATVTRQDEGTYSRVEDSVDFRLVSPKVPVFAKEQLARRKL 900
841 GERFLOEDVGTGTHRLVAATVTRQDEGTYSRVEDSVDFRLVSPKVPVFAKEQLARRKL 900
901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVCMKMEATGCTRRLLVQQAGQADAGEY 960
901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVCMKMEATGCTRRLLVQQAGQADAGEY 960
961 SCEAGGQRLSFHLDVKEPKVFAKQVAHSEYQAEAGANATLSCEVAQAQAEVVMYKDGK 1020
961 SCEAGGQRLSFHLDVKEPKVFAKQVAHSEYQAEAGANATLSCEVAQAQAEVVMYKDGK 1020
1021 KLSSSSKVHVEAKGCRRLVVOAGKTADGYSCEARGORVSRFRHLHITPEKVMFAKEQSV 1080
1021 KLSSSSKVHVEAKGCRRLVVOAGKTADGYSCEARGORVSRFRHLHITPEKVMFAKEQSV 1080
1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSSKVGMVEKGTERRLLVPOAGKAD 1140
1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSSKVGMVEKGTERRLLVPOAGKAD 1140
1141 AGEYSECEAGGQRVSRFHLHITPEKGVFAKEQSVHNEVQAEAGTAMLSCEVAQOPQTEVTWY 1200
1141 AGEYSECEAGGQRVSRFHLHITPEKGVFAKEQSVHNEVQAEAGTAMLSCEVAQOPQTEVTWY 1200
1201 KDGKLSKSSSKVMEVKGCTRRLLVVOAGKADAGEYSCGAGQORVSRFHLHITPEKAVPAK 1260
1201 KDGKLSKSSSKVMEVKGCTRRLLVVOAGKADAGEYSCGAGQORVSRFHLHITPEKAVPAK 1260
1261 EQLVHNEVTEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRIEAGCQRLVVOQA 1320
1261 EQLVHNEVTEAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVRIEAGCQRLVVOQA 1320
1321 GQADAGEYTCGAGGQRLSFHLDVSEPKAVFAKEQLAHRKVQAEAGATLSCEVAQAQTE 1380
1321 GQADAGEYTCGAGGQRLSFHLDVSEPKAVFAKEQLAHRKVQAEAGATLSCEVAQAQTE 1380
1381 VTWYKDGKLSKSSSKVMEVKGCTRRLLVVOQAQADTGYSCBAGQORLSFSLDVABPKV 1440
1381 VTWYKDGKLSKSSSKVMEVKGCTRRLLVVOQAQADTGYSCBAGQORLSFSLDVABPKV 1440
1441 VFAKEQPVHREVOQAQAGATLSCEVAQAQTEVTWYKDGKLSKSSSKVMEVKGCTRRLLV 1500
1441 VFAKEQPVHREVOQAQAGATLSCEVAQAQTEVTWYKDGKLSKSSSKVMEVKGCTRRLLV 1500
1501 VQAGQADAGEYSCGAGGQRLSFHLHVAEPKAVFAKEQPAEVEVQAEAGTATLSCEVAQ 1560
1501 VQAGQADAGEYSCGAGGQRLSFHLHVAEPKAVFAKEQPAEVEVQAEAGTATLSCEVAQ 1560
1561 AQTEVTWYKDGKLSKSSSKVMEVKGCTRRLLVVOEAGQADAGEYSCGAGQORLSFHLHVA 1620
1561 AQTEVTWYKDGKLSKSSSKVMEVKGCTRRLLVVOEAGQADAGEYSCGAGQORLSFHLHVA 1620
1621 EPKVPFAKEQPAHREVOQAAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVMEVKGCT 1680
1621 EPKVPFAKEQPAHREVOQAAGASATLSCEVAQAQTEVTWYKDGKLSKSSSKVMEVKGCT 1680
1681 RRLVVOQAQADAGEYSCGAGGQRLSFHLHVAEPKAVFAKEQPAEVEVQAEAGTATLSCEVAQ 1740

```

1681 RRLVVCQAGADAGEYSCEAGQRLSFLHVAEPEQISERPCREPLVWKEHDIILTA 1740
1741 TLATPSAATVTLKDGVEIRRSKRHETASQGDTHTLVHGAQVLDLSAISYCRVGAEGQDF 1800
1741 TLATPSAATVTLKDGVEIRRSKRHETASQGDTHTLVHGAQVLDLSAISYCRVGAEGQDF 1800
1801 PVQVEEVAAKTCLLEPVCGBLGGTVTLACBELSPACAEVVMRCNTQPRVKGKPFQWVAEG 1860
1801 PVQVEEVAAKTCLLEPVCGBLGGTVTLACBELSPACAEVVMRCNTQPRVKGKPFQWVAEG 1860
1861 PVRSLTVLGLRAEDAGEVYCCSRDDHTSAQLTVSPVVKFMSGLSTVVAEGEATFQC 1920
1861 PVRSLTVLGLRAEDAGEVYCCSRDDHTSAQLTVSPVVKFMSGLSTVVAEGEATFQC 1920
1921 VVSPSDVAVWFRDGLLOPSEKFAISQSGASHSLTISDLVLEDAGQITVVEAGASSAA 1980
1921 VVSPSDVAVWFRDGLLOPSEKFAISQSGASHSLTISDLVLEDAGQITVVEAGASSAA 1980
1981 LRVRAPVLFKKLEPQTVBERSSTVLELTPRMPBLRWTRNATALAGKNVEIHAEGA 2040
1981 LRVRAPVLFKKLEPQTVBERSSTVLELTPRMPBLRWTRNATALAGKNVEIHAEGA 2040
2041 RRLVHLNVGADRGFFGFCETPDDKTQAKLTVMRQVRLVRGLQAVEAREQGTATMEVOL 2100
2041 RRLVHLNVGADRGFFGFCETPDDKTQAKLTVMRQVRLVRGLQAVEAREQGTATMEVOL 2100
2101 SHADVGSWTRDGLRFQGPPTCHLAVRGPMTHTLSGLRPEDSGLVWFKAEGVHTSARLV 2160
2101 SHADVGSWTRDGLRFQGPPTCHLAVRGPMTHTLSGLRPEDSGLVWFKAEGVHTSARLV 2160
2161 VTLPVPSRPLQDVVTEKEKVTLECELSRPVNDVWELKDGVELRACKTMAAAGACR 2220
2221 SLTIYRCBFDAGQVYVCDADAQASASVKGQRTYTLIYRRVLAEDAGEIQFVAENAESR 2280
2221 SLTIYRCBFDAGQVYVCDADAQASASVKGQRTYTLIYRRVLAEDAGEIQFVAENAESR 2280
2281 AOLRVKELPVTILRPLRDKTAMEKHGVLCEQVSRASAOVRPKGSOELQPGKYELVSD 2340
2281 AOLRVKELPVTILRPLRDKTAMEKHGVLCEQVSRASAOVRPKGSOELQPGKYELVSD 2340
2341 GLYKRLIISDVHAEDEDTYCDAGDVKTSQFFVVEQSIITVIRGLQDVVMEPAPAFEC 2400
2341 GLYKRLIISDVHAEDEDTYCDAGDVKTSQFFVVEQSIITVIRGLQDVVMEPAPAFEC 2400
2401 ETSIPSVRPPKWLGLKTVLQAGNVGLQESGTVHRLMLRRTCTMTGCPVHFTVKGSRSSA 2460
2401 ETSIPSVRPPKWLGLKTVLQAGNVGLQESGTVHRLMLRRTCTMTGCPVHFTVKGSRSSA 2460
2461 RLVSVDIPVLTREPLEKTRGLQSVVLSCDFPAPKAVQWYKDDTPLSPSEKFKVSLLEG 2520
2461 RLVSVDIPVLTREPLEKTRGLQSVVLSCDFPAPKAVQWYKDDTPLSPSEKFKVSLLEG 2520
2521 QMAELRILRLMPADAGVYRCQASAHSSSTVEAREVTVTGLODABATEBWASFSCE 2580
2521 QMAELRILRLMPADAGVYRCQASAHSSSTVEAREVTVTGLODABATEBWASFSCE 2580
2581 LSHDEEVEVSLNGMPLYNDSFHEISHKGRHRTLVLKSIQADAGIVRASSLKVSTARSIL 2640
2581 LSHDEEVEVSLNGMPLYNDSFHEISHKGRHRTLVLKSIQADAGIVRASSLKVSTARSIL 2640
2641 EVRKVPVFTKALDDLSAEBRGTLALQCEVSDPEAHVVMKQVQLGSPDKYDFLHTAGT 2700
2641 EVRKVPVFTKALDDLSAEBRGTLALQCEVSDPEAHVVMKQVQLGSPDKYDFLHTAGT 2700
2701 RGLVHVDVSPEDAGLYTCHVGSSETRARVRVHDLHVGITKRLKTMVELEGESCFECVL 2760
2701 RGLVHVDVSPEDAGLYTCHVGSSETRARVRVHDLHVGITKRLKTMVELEGESCFECVL 2760
2761 HESASDPAMWTVGGKTVGSSSRFOATROGRKYILVWREAAPSDAGEVVFVSRGLTSKASL 2820

2761 HESASDPAMWTVGGKTVGSSSRFOATROGRKYILVWREAAPSDAGEVVFVSRGLTSKASL 2820
2821 IYRERPAALIKPLEDQWVAPGEDVELRCELGRAGTFVHMLKDRKAIRKSQKYDVVCEGTM 2880
2821 IYRERPAALIKPLEDQWVAPGEDVELRCELGRAGTFVHMLKDRKAIRKSQKYDVVCEGTM 2880
2881 AMLVIRGASLKDAGBYTCEVEASKSTASLHVEEKANCFTBELTNLOVEKGTAVFTCKTE 2940
2881 AMLVIRGASLKDAGBYTCEVEASKSTASLHVEEKANCFTBELTNLOVEKGTAVFTCKTE 2940
2941 HFAAATVWRKGLLELRASGKHQPSOEGTLRLTISALEKADSDTTCIDIGOAQRAQLLV 3000
2941 HFAAATVWRKGLLELRASGKHQPSOEGTLRLTISALEKADSDTTCIDIGOAQRAQLLV 3000
3001 QGRRVHIIIEDLVVDVQEGSSATFRCRISPANYEPVHFWLTKTPLHANELNEIDAQPGY 3060
3001 QGRRVHIIIEDLVVDVQEGSSATFRCRISPANYEPVHFWLTKTPLHANELNEIDAQPGY 3060
3061 HVLTLRQALKDSGTIYFEAGDQASAAALRVTEKPSVPSRELTDATITTEGEDTLVCEFS 3120
3061 HVLTLRQALKDSGTIYFEAGDQASAAALRVTEKPSVPSRELTDATITTEGEDTLVCEFS 3120
3121 TCDIPMCWTQDKGKTLRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGCACSSIVRV 3180
3121 TCDIPMCWTQDKGKTLRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGCACSSIVRV 3180
3181 HARPVRFQALKDLLEVLEGGATLRVLSVAAPVKKCYGNVLRPGDKYSLRQEGAMLE 3240
3181 HARPVRFQALKDLLEVLEGGATLRVLSVAAPVKKCYGNVLRPGDKYSLRQEGAMLE 3240
3241 LVVRNLRQDSGRYSCSPGDTTISATLTVTALPAQFICKLENKEATGATATLCELSKT 3300
3241 LVVRNLRQDSGRYSCSPGDTTISATLTVTALPAQFICKLENKEATGATATLCELSKT 3300
3301 APVEWRKSETLRQDGRVCLRQDGMCELOIRGLAMVDAABYSVCVEERTSASLTTRPM 3360
3301 APVEWRKSETLRQDGRVCLRQDGMCELOIRGLAMVDAABYSVCVEERTSASLTTRPM 3360
3361 PAHFTGRURHOESLEGATATLCELKSAAPVEWRKGRSLRQDGRHSLRQDGAQCELOIC 3420
3361 PAHFTGRURHOESLEGATATLCELKSAAPVEWRKGRSLRQDGRHSLRQDGAQCELOIC 3420
3421 GLAVADAGEYSCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAVLMCELSKVAPVE 3480
3421 GLAVADAGEYSCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAVLMCELSKVAPVE 3480
3481 WRKGFENLRDGRYTLRQEGTRCELOICGLAMADAGEVLCVCGOERTSATLITRALPARF 3540
3481 WRKGFENLRDGRYTLRQEGTRCELOICGLAMADAGEVLCVCGOERTSATLITRALPARF 3540
3541 IEDVKNQOAREGATAVLQCELSAAPVEWRKGSSETLRDGRYSLRQDGTKCELOIRGLAM 3600
3541 IEDVKNQOAREGATAVLQCELSAAPVEWRKGSSETLRDGRYSLRQDGTKCELOIRGLAM 3600
3601 ADTGEYSVCVCGOERTSATLTVRALPIKFTTEGLRNEEAVEGATAVLMCELSKVAPVEWKG 3660
3601 ADTGEYSVCVCGOERTSATLTVRALPIKFTTEGLRNEEAVEGATAVLMCELSKVAPVEWKG 3660
3661 HETLRDGRHSLRQDGRARCELQIRGLVAEDAGEVLCVCGOERTSATLTVRAMPSKFIEGL 3720
3661 HETLRDGRHSLRQDGRARCELQIRGLVAEDAGEVLCVCGOERTSATLTVRAMPSKFIEGL 3720
3721 RNEATEGDTATLMCELSKAAAPVEWRKGHETLRDGRHSLRQDGRARCELQIRGLAVVDAG 3780
3721 RNEATEGDTATLMCELSKAAAPVEWRKGHETLRDGRHSLRQDGRARCELQIRGLAVVDAG 3780
3781 EYSCVCGOERTSATLTVRALPARFIEDVKNQOAREGATAVLQCELSKAAAPVEWRKGSSETL 3840
3781 EYSCVCGOERTSATLTVRALPARFIEDVKNQOAREGATAVLQCELSKAAAPVEWRKGSSETL 3840
3841 RGGDRYSLRQDGTARCELQIHGLSVADTGEYSVCVCGOERTSATLTVRAPQVFPREPLOSQ 3900
3841 RGGDRYSLRQDGTARCELQIHGLSVADTGEYSVCVCGOERTSATLTVRAPQVFPREPLOSQ 3900

| | | | |
|----|------|--|------|
| QY | 3901 | ABEGSTATLQCELSBPTATVWWSKGLQLOANGREPREPLOGCTABELVLQDLOREBDTGEYT | 3960 |
| Db | 3901 | ABEGSTATLQCELSBPTATVWWSKGLQLOANGREPREPLOGCTABELVLQDLOREBDTGEYT | 3960 |
| QY | 3961 | CTCGSOATATVTTAAVPVRFRELOHQHVEDEGGTAHLCCELSRAGASVEMRKSLQLFP | 4020 |
| Db | 3961 | CTCGSOATATVTTAAVPVRFRELOHQHVEDEGGTAHLCCELSRAGASVEMRKSLQLFP | 4020 |
| QY | 4021 | CAKYQWQDGAALLLVGVQEDAGDYTCDTGHTQSMASLSVRPRPKPTRLOSLEQE | 4080 |
| Db | 4021 | CAKYQWQDGAALLLVGVQEDAGDYTCDTGHTQSMASLSVRPRPKPTRLOSLEQE | 4080 |
| QY | 4081 | TGDIARLCCOLSDAESGAVQWVKEGVELHAGPKYEMESQCATRELLIHQLEAKDTGEYA | 4140 |
| Db | 4081 | TGDIARLCCOLSDAESGAVQWVKEGVELHAGPKYEMESQCATRELLIHQLEAKDTGEYA | 4140 |
| QY | 4141 | CVTGQOKTAASLRVTEPEVTIVRGLVDAEVTADEDFEFSCEVSRAGATGVQWCLQGLPLQ | 4200 |
| Db | 4141 | CVTGQOKTAASLRVTEPEVTIVRGLVDAEVTADEDFEFSCEVSRAGATGVQWCLQGLPLQ | 4200 |
| QY | 4201 | SNEVTEAVRDRGRIHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVITILEPQDVQ | 4260 |
| Db | 4201 | SNEVTEAVRDRGRIHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVITILEPQDVQ | 4260 |
| QY | 4261 | LSEGODASFQCLSRASQOEARWALGGVPLQANEMNDITVBQGTLLHLLTLHKVTLDEAGT | 4320 |
| Db | 4261 | LSEGODASFQCLSRASQOEARWALGGVPLQANEMNDITVBQGTLLHLLTLHKVTLDEAGT | 4320 |
| QY | 4321 | VSFHVGTCSSEALQKVTAKNTVVRGLENVLEALEGGEALFEQOLSQPEVAANTWLLDDEPV | 4380 |
| Db | 4321 | VSFHVGTCSSEALQKVTAKNTVVRGLENVLEALEGGEALFEQOLSQPEVAANTWLLDDEPV | 4380 |
| QY | 4381 | RTSENAEVVFPENGRLHLLLNLRPQDSRCVTFFLAGDMVTSAPLTVRGWRLTILEPLKN | 4440 |
| Db | 4381 | RTSENAEVVFPENGRLHLLLNLRPQDSRCVTFFLAGDMVTSAPLTVRGWRLTILEPLKN | 4440 |
| QY | 4441 | AAVRAGAQAFTCTLSEAVPVGEASWYINGAAVQPDSDMTVTADGSHQALLLSAQPHH | 4500 |
| Db | 4441 | AAVRAGAQAFTCTLSEAVPVGEASWYINGAAVQPDSDMTVTADGSHQALLLSAQPHH | 4500 |
| QY | 4501 | AGEVTFACRDVAVASARLTVILGLPDPPEDAEVVAHSSHTVTLSPAAPMSDGGGGICGVRYE | 4560 |
| Db | 4501 | AGEVTFACRDVAVASARLTVILGLPDPPEDAEVVAHSSHTVTLSPAAPMSDGGGGICGVRYE | 4560 |
| QY | 4561 | VKEGATGWRLLCHELVPGPECVVDGLAPGETYRFRVAAVGPVGAGEPVHLPTQTVRLAEP | 4620 |
| Db | 4561 | VKEGATGWRLLCHELVPGPECVVDGLAPGETYRFRVAAVGPVGAGEPVHLPTQTVRLAEP | 4620 |
| QY | 4621 | KVPVQPSAPSRQVAAAGEDVSLELVVAEAGEVIVHKGMBRIQPGGRFEVVSQGRQML | 4680 |
| Db | 4621 | KVPVQPSAPSRQVAAAGEDVSLELVVAEAGEVIVHKGMBRIQPGGRFEVVSQGRQML | 4680 |
| QY | 4681 | VIKGTAFDQGEYHCGLAQSGICPAAATFQVALSPASVDEAPQPSLPPPEAAQEGDLHLW | 4740 |
| Db | 4681 | VIKGTAFDQGEYHCGLAQSGICPAAATFQVALSPASVDEAPQPSLPPPEAAQEGDLHLW | 4740 |
| QY | 4741 | EALAKRMSREPTLDSISLPEEDGRSQRILPQBAEEVAPDLSEGYSTADELARTGDADL | 4800 |
| Db | 4741 | EALAKRMSREPTLDSISLPEEDGRSQRILPQBAEEVAPDLSEGYSTADELARTGDADL | 4800 |
| QY | 4801 | SHTSSDDRESRAGTPSLVTLKAGBPGTSLASKVGAPAAPSVKPEQOOEPLAARPLG | 4860 |
| Db | 4801 | SHTSSDDRESRAGTPSLVTLKAGBPGTSLASKVGAPAAPSVKPEQOOEPLAARPLG | 4860 |
| QY | 4861 | DLSTKDLDGPPSKDAAVKIQAAFKGYKVRKEMKQOEGPMSFHTFGDTEAQVGDALRLECV | 4920 |
| Db | 4861 | DLSTKDLDGPPSKDAAVKIQAAFKGYKVRKEMKQOEGPMSFHTFGDTEAQVGDALRLECV | 4920 |
| QY | 4921 | VASKADVBARMLKQGVETLDGRHHHIDOLGDBGTCSLLIAGLDRADAGCYTCQVSNKFGQV | 4980 |
| Db | 4921 | VASKADVBARMLKQGVETLDGRHHHIDOLGDBGTCSLLIAGLDRADAGCYTCQVSNKFGQV | 4980 |
| QY | 4981 | THSACVVVSGSSEAEASSSGGELDDAFRAARARLHLFRITKSPAESVDEBELFLSABEGPA | 5040 |
| Db | 4981 | THSACVVVSGSSEAEASSSGGELDDAFRAARARLHLFRITKSPAESVDEBELFLSABEGPA | 5040 |
| QY | 5041 | EPPEADWQYRDEDEHFCIRFEALTEARQAVTRFOEMFATLIGIVIEIKLVEOGPRVEM | 5100 |
| Db | 5041 | EPPEADWQYRDEDEHFCIRFEALTEARQAVTRFOEMFATLIGIVIEIKLVEOGPRVEM | 5100 |
| QY | 5101 | CISKETPAPVVPPEPILSLTSDAAPVFLTELOQVQDQGVPSFDCVVTGQPMPSVRWF | 5160 |
| Db | 5101 | CISKETPAPVVPPEPILSLTSDAAPVFLTELOQVQDQGVPSFDCVVTGQPMPSVRWF | 5160 |
| QY | 5161 | KDKKLEEDDHVMINEDQGGHOLIITAVVPADMVYRCLAENSMGVSSTKAERLVDLTS | 5220 |
| Db | 5161 | KDKKLEEDDHVMINEDQGGHOLIITAVVPADMVYRCLAENSMGVSSTKAERLVDLTS | 5220 |
| QY | 5221 | TDYDTAADATESSESYFSAQYLSRRQEGTSTTDEGQLPVVVEELDLQVAPGTRLAKF | 5280 |
| Db | 5221 | TDYDTAADATESSESYFSAQYLSRRQEGTSTTDEGQLPVVVEELDLQVAPGTRLAKF | 5280 |
| QY | 5281 | OLKVGYPAARLYWFKDGOPLTASAHIRMTGKILHTLBIISVTREDSQYAAIYNAMG | 5340 |
| Db | 5281 | OLKVGYPAARLYWFKDGOPLTASAHIRMTGKILHTLBIISVTREDSQYAAIYNAMG | 5340 |
| QY | 5341 | AAYSARLLVRGPDPEPEKXPASDVHEQLVPPRMLERFTPKVKYKGSISITFSVKEGRPVP | 5400 |
| Db | 5341 | AAYSARLLVRGPDPEPEKXPASDVHEQLVPPRMLERFTPKVKYKGSISITFSVKEGRPVP | 5400 |
| QY | 5401 | TVHMLREABRGVLTGPDTPGYTVASSAQCHSLVLLDVGRQHOGTYTCIASNAQALC | 5460 |
| Db | 5401 | TVHMLREABRGVLTGPDTPGYTVASSAQCHSLVLLDVGRQHOGTYTCIASNAQALC | 5460 |
| QY | 5461 | SASLHVSGLPKVBEQKVEALISTFLOGTTOAISAQLETFADLGGORKEEPLAAKE | 5520 |
| Db | 5461 | SASLHVSGLPKVBEQKVEALISTFLOGTTOAISAQLETFADLGGORKEEPLAAKE | 5520 |
| QY | 5521 | ALGHLSLAEVGTBEFTLQKLTQITTEMVSAKIIOAKLVQPGGSDSDSKTPSPSPRGRSR | 5580 |
| Db | 5521 | ALGHLSLAEVGTBEFTLQKLTQITTEMVSAKIIOAKLVQPGGSDSDSKTPSPSPRGRSR | 5580 |
| QY | 5581 | PSSSIOESSESSEDGARGEIDFIIYVVTADYLPGRQDQAITLREGQYVEVLDAAHPLW | 5640 |
| Db | 5581 | PSSSIOESSESSEDGARGEIDFIIYVVTADYLPGRQDQAITLREGQYVEVLDAAHPLW | 5640 |
| QY | 5641 | LVRTKTKSPSRQGVWSPAYLDRRLKLSPEWGAAPFPPEGAESVDESEYKARLSVIOE | 5700 |
| Db | 5641 | LVRTKTKSPSRQGVWSPAYLDRRLKLSPEWGAAPFPPEGAESVDESEYKARLSVIOE | 5700 |
| QY | 5701 | LLSSEQAUFVEELOFLQSHHLOHLERCPHVPIAVAGOKAVIFRNRDIGHFHSFLQELQ | 5760 |
| Db | 5701 | LLSSEQAUFVEELOFLQSHHLOHLERCPHVPIAVAGOKAVIFRNRDIGHFHSFLQELQ | 5760 |
| QY | 5761 | CDTDDDVAMCFIKNQAAFEQYLEFLVGRVOAESVVVSTAIQBFYKKYAEALLAGDPSQ | 5820 |
| Db | 5761 | CDTDDDVAMCFIKNQAAFEQYLEFLVGRVOAESVVVSTAIQBFYKKYAEALLAGDPSQ | 5820 |
| QY | 5821 | PPPLQHYLEQPVVERVQRYQALLKELIRKARNRQNCALLEQAYAVVSALPQRAENKLHV | 5880 |
| Db | 5821 | PPPLQHYLEQPVVERVQRYQALLKELIRKARNRQNCALLEQAYAVVSALPQRAENKLHV | 5880 |
| QY | 5881 | SLMENYPGTLEALGEPDROGHFIVWEGAPCARPKGNHNVFLFNHJVIKCPREDSRT | 5940 |
| Db | 5881 | SLMENYPGTLEALGEPDROGHFIVWEGAPCARPKGNHNVFLFNHJVIKCPREDSRT | 5940 |
| QY | 5941 | DTVSYVFRNMKLLSIDLNDQVEGDDRAFEVWQEREDSVRKYLQARTAIKSSWKEIC | 6000 |
| Db | 5941 | DTVSYVFRNMKLLSIDLNDQVEGDDRAFEVWQEREDSVRKYLQARTAIKSSWKEIC | 6000 |
| QY | 6001 | GIQORLALPWRPPDPEEBLADCTABGETVKLACHVTGTPKPVTSWYKDGKAVQVDPH | 6060 |
| Db | 6001 | GIQORLALPWRPPDPEEBLADCTABGETVKLACHVTGTPKPVTSWYKDGKAVQVDPH | 6060 |
| QY | 6061 | ILIEDPDGSCALITLDTGVDGQVCFMFAASAGNCSITLGLKILVQVPRPFVNVKVRASPFV | 6120 |

```
Db 6061 ILIEDPGSCALILDSLTGVDSGQYMCFAASAAGNCSTLKGILVQPPRVNKLSPFV 6120
Qy 6121 EGEDAQFTCTTEGAPYQIRWKDGLTTGNKFTLSEPRSGLLVLVIRAAKEDLGLY 6180
Db 6121 EGEDAQFTCTTEGAPYQIRWKDGLTTGNKFTLSEPRSGLLVLVIRAAKEDLGLY 6180
Qy 6181 BCELVNRLGSRASAEIRIQSPMLQAOEQCHREQLVAIVEDTTLER-----ADQEV 6232
Db 6181 BCELVNRLGSRASAEIRIQSPMLQAOEQCHREQLVAIVEDTTLER-----ADQEV 6232
Qy 6233 SVLXRLGLKAPGSTGDLTGPGCPRG-----APAL-----QETSQP- 6271
Db 6241 TVKSPRQRR-SPKSPRSRSCSASPLRPLGLAPDLYLPGAGOPRPEAPGQKPV 6299
Qy 6272 -PVTGTSEAPVPRV-----PQLLHEGPE-----QPEAIARAQEWTPIRMEGA 6317
Db 6300 VPTLYVTEAEASHPALGSLGPGPKWVEETIEVRVKXGPGQVSPTE--VPRSSGH 6357
Qy 6318 AW--PGACTG 6325
Db 6358 LFTLPGATPG 6367

RESULT 4
US-10-077-130-2
; Sequence 2, Application US/10077130
; Publication No. US20020169742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Thereof
; FILE REFERENCE: MPI2001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2630
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-077-130-2

Query Match 33.2%; Score 13710; DB 13; Length 2630;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2630; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5339 MGAAYSSARLLVGPDEPEEKPSADVHEQLVPPRMLEFFTPKVKKGSSITFSVKVEGRP 5398
Db 1 MGAAYSSARLLVGPDEPEEKPSADVHEQLVPPRMLEFFTPKVKKGSSITFSVKVEGRP 60

Qy 5399 VPTVHMLREEAERGLVMIPTDPTGYTVASSAQHSLVLLDVGRHQHGYTCIASNAAGQA 5458
Db 61 VPTVHMLREEAERGLVMIPTDPTGYTVASSAQHSLVLLDVGRHQHGYTCIASNAAGQA 120

Qy 5459 LCSASLVHSGLPKVEEKEKVEALISTFLOTTQAIISAQGLETFASFDLGGQKEEPLAA 5518
Db 121 LCSASLVHSGLPKVEEKEKVEALISTFLOTTQAIISAQGLETFASFDLGGQKEEPLAA 180

Qy 5519 KEALCHLSLAVGTEEFLOKLTSTOITEMWSAKITQAKLVPGGSDSDSKTPASPRHGR 5578
Db 181 KEALCHLSLAVGTEEFLOKLTSTOITEMWSAKITQAKLVPGGSDSDSKTPASPRHGR 240

Qy 5579 SRPSSSQESSSESDGDARGEIFDIYVVTADYPLGAEOQATLREGQVVEVLDAAHPL 5638
Db 241 SRPSSSQESSSESDGDARGEIFDIYVVTADYPLGAEOQATLREGQVVEVLDAAHPL 300

Qy 5639 RNLVTRTKSSPSRQGVSPAYLDRRLKLSPEWGAAEAPEFFGCEAVSEDEYKARLSSVI 5698
Db 301 RNLVTRTKSSPSRQGVSPAYLDRRLKLSPEWGAAEAPEFFGCEAVSEDEYKARLSSVI 360
```

```
Qy 5699 QELLSEQAFVEEQLOFLOSHHLQHLERCPHPVPIAVAGQKAVIFRNVRDIGRHFHSFLQEL 5758
Db 361 QELLSEQAFVEEQLOFLOSHHLQHLERCPHPVPIAVAGQKAVIFRNVRDIGRHFHSFLQEL 420
Qy 5759 QQCDDDDVAMCFIKNQAQFEQYLEFLVGRVOAESVVVSTAIQBFYKKYABEALLAGDPS 5818
Db 421 QQCDDDDVAMCFIKNQAQFEQYLEFLVGRVOAESVVVSTAIQBFYKKYABEALLAGDPS 480
Qy 5819 QPPPPLOHYLEQPVVERVORYOALLKELIRKANRQNCALLEQAYAVVSALPORAENKL 5878
Db 481 QPPPPLOHYLEQPVVERVORYOALLKELIRKANRQNCALLEQAYAVVSALPORAENKL 540
Qy 5879 HVSLMENYPGTLEALGEPIRQGHFIWEGAPGARMKGNHNRHVLFRNHLVICRPRDS 5938
Db 541 HVSLMENYPGTLEALGEPIRQGHFIWEGAPGARMKGNHNRHVLFRNHLVICRPRDS 600
Qy 5939 RTDVTYSYFRNMKLSIDLDNDQVEGDDRAFEVWQEREDSVRKYLLQARTAIKSSWYKE 5998
Db 601 RTDVTYSYFRNMKLSIDLDNDQVEGDDRAFEVWQEREDSVRKYLLQARTAIKSSWYKE 660
Qy 5999 ICGIQORLALPVWRPDPPEELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDP 6058
Db 661 ICGIQORLALPVWRPDPPEELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDP 720
Qy 6059 KHILIEDPGSCALILDSLTGVDSGQYMCFAASAAGNCSTLKGILVQPPRVNKLSPFV 6118
Db 721 KHILIEDPGSCALILDSLTGVDSGQYMCFAASAAGNCSTLKGILVQPPRVNKLSPFV 780
Qy 6119 FVEGEDAQFTCTTEGAPYQIRWKDGLTTGNKFTLSEPRSGLLVLVIRAAKEDLGL 6178
Db 781 FVEGEDAQFTCTTEGAPYQIRWKDGLTTGNKFTLSEPRSGLLVLVIRAAKEDLGL 840
Qy 6179 LYECELVNRLGSRASAEIRIQSPMLQAOEQCHREQLVAIVEDTTLERADQEVTSVLKRL 6238
Db 841 LYECELVNRLGSRASAEIRIQSPMLQAOEQCHREQLVAIVEDTTLERADQEVTSVLKRL 900
Qy 6239 LGPKAPGSTGDLTGPGCPRGAPALQETGSPVTCGTSEAPVPPVPPQPLLHEGPEQE 6298
Db 901 LGPKAPGSTGDLTGPGCPRGAPALQETGSPVTCGTSEAPVPPVPPQPLLHEGPEQE 960
Qy 6299 PEAIARAQEWTPVIRMEGAAMPAGCTGELLMDVHSHVVRETTQRTTYQAIDTHTARPPS 6358
Db 961 PEAIARAQEWTPVIRMEGAAMPAGCTGELLMDVHSHVVRETTQRTTYQAIDTHTARPPS 1020
Qy 6359 MQVTIEDVQAQGTGTAQFEALIEGDPQPSVTWYKDSVQLVDSTSLSCQOEGTTSVLVRH 6418
Db 1021 MQVTIEDVQAQGTGTAQFEALIEGDPQPSVTWYKDSVQLVDSTSLSCQOEGTTSVLVRH 1080
Qy 6419 VASKDAGVYTCLAQNTGQVLCKAELLVLGGDNEPDSEKQSHRRKLHSFYEVKEEIGRGV 6478
Db 1081 VASKDAGVYTCLAQNTGQVLCKAELLVLGGDNEPDSEKQSHRRKLHSFYEVKEEIGRGV 1140
Qy 6479 FGFVKRVQHKNGKILCAAKFIPLRSRTRAQYRREDILAAHLSHPVLTGLLDQFETRTKLI 6538
Db 1141 FGFVKRVQHKNGKILCAAKFIPLRSRTRAQYRREDILAAHLSHPVLTGLLDQFETRTKLI 1200
Qy 6539 LLELCSSSEELLDRLYRGVVTAEVKYIQQLEGLHYLHSHGVHLHDIKPSNLMVHP 6598
Db 1201 LLELCSSSEELLDRLYRGVVTAEVKYIQQLEGLHYLHSHGVHLHDIKPSNLMVHP 1260
Qy 6599 AREDIKICDFGAQNTIPAEILQFSGYSGSPFVSPEIQQONPVSEASDIWANGVTSYLSLT 6658
Db 1261 AREDIKICDFGAQNTIPAEILQFSGYSGSPFVSPEIQQONPVSEASDIWANGVTSYLSLT 1320
Qy 6659 CSSPAGESDRATLNVLEGRVSWSSPMAHLSDADDFIKATLQAPQAPPSAQCLSH 6718
Db 1321 CSSPAGESDRATLNVLEGRVSWSSPMAHLSDADDFIKATLQAPQAPPSAQCLSH 1380
Qy 6719 PMFLKMPAEAEHAFINTKQLKFLARSQRWSLMSYKSIYMRSPPELLRGPPDPSGLV 6778
Db 1381 PMFLKMPAEAEHAFINTKQLKFLARSQRWSLMSYKSIYMRSPPELLRGPPDPSGLV 1440
```

QY 6779 ASHLCDTGGSSSSSSSDNELAPFARAKSLPPSPVTHSPHLLHPRGFLRPSASIPERAEA 6838
DB 1441 ASHLCDTGGSSSSSSSDNELAPFARAKSLPPSPVTHSPHLLHPRGFLRPSASIPERAEA 1500
QY 6839 SERSTEAPAPASPGAGAPPAAGCVPRHVSIRSLFYHQAGSEPHGALAPGSRHHPARR 6898
DB 1501 SERSTEAPAPASPGAGAPPAAGCVPRHVSIRSLFYHQAGSEPHGALAPGSRHHPARR 1560
QY 6899 RHLKGGYIAGALPGLRPLMEHRVLEBEAREOATLLAKAPFETALRPLPASGTHLAP 6958
DB 1561 RHLKGGYIAGALPGLRPLMEHRVLEBEAREOATLLAKAPFETALRPLPASGTHLAP 1620
QY 6959 GHSHLEHDSSTPRPSEACEAQLPSAPSGGAPIDMGHPGSGKOLPSTGCHPGTAQ 7018
DB 1621 GHSHLEHDSSTPRPSEACEAQLPSAPSGGAPIDMGHPGSGKOLPSTGCHPGTAQ 1680
QY 7019 PERPSPDWQOPAPFCHPKQGSAPQECSPHPAVAPCPGSPFPGCKEAPLVPSPPFL 7078
DB 1681 PERPSPDWQOPAPFCHPKQGSAPQECSPHPAVAPCPGSPFPGCKEAPLVPSPPFL 1740
QY 7079 GQPAPPAKASPLDCKMGPGLISLGRPKGPCSPGSGASQSSQVSSLRVGSQV 7138
DB 1741 GQPAPPAKASPLDCKMGPGLISLGRPKGPCSPGSGASQSSQVSSLRVGSQV 1800
QY 7139 GTEPGPSLDAEGWTQEAEDLSNSTPTLQRPQEQATMRKFSLGGRGYAGVAGYGTFAFGG 7198
DB 1801 GTEPGPSLDAEGWTQEAEDLSNSTPTLQRPQEQATMRKFSLGGRGYAGVAGYGTFAFGG 1860
QY 7199 DAGMLOGPMMWARTAMAVSQSEBEQEAERAEQSEBQQAARAESELPQVARSAPVEVG 7258
DB 1861 DAGMLOGPMMWARTAMAVSQSEBEQEAERAEQSEBQQAARAESELPQVARSAPVEVG 1920
QY 7259 RAPTRSSPEPTWEDIGQVSLVQIRDLSDGDEADTTSLDISEVDPAVLNLSLDLYDKYL 7318
DB 1921 RAPTRSSPEPTWEDIGQVSLVQIRDLSDGDEADTTSLDISEVDPAVLNLSLDLYDKYL 1980
QY 7319 PFEMIFPKVKPSAQPEPPSPWABEELAEFFPEPTWMPGELGPHAGLEITESESDVDALL 7378
DB 1981 PFEMIFPKVKPSAQPEPPSPWABEELAEFFPEPTWMPGELGPHAGLEITESESDVDALL 2040
QY 7379 AEAAGRKRKWSRSRSLFHPGPHLPLDEPAEGLERBKVASVEHLSRLKGRPEGLEK 7438
DB 2041 AEAAGRKRKWSRSRSLFHPGPHLPLDEPAEGLERBKVASVEHLSRLKGRPEGLEK 2100
QY 7439 EGPRKPKGLASFRISGLKSDWDRAPTFRLRLSDETVLQSVTLACQVSAQPAQAATWSK 7498
DB 2101 EGPRKPKGLASFRISGLKSDWDRAPTFRLRLSDETVLQSVTLACQVSAQPAQAATWSK 2160
QY 7499 DGAPLESSRVLISATLKNFOLLILVVAEDLGYITCSVSNALGTVTITGVLKARPS 7558
DB 2161 DGAPLESSRVLISATLKNFOLLILVVAEDLGYITCSVSNALGTVTITGVLKARPS 2220
QY 7559 SSPCPDICEVADGVLVVKPVESYGPVTYIVQCSLEGGSWTTLASDIFDCCYLTSKLSR 7618
DB 2221 SSPCPDICEVADGVLVVKPVESYGPVTYIVQCSLEGGSWTTLASDIFDCCYLTSKLSR 2280
QY 7619 GGTYTFRTACVSKAGMGYSSPSEQVLLGGPSHLASEEESQGRSAQPLPSTKTAFOTQI 7678
DB 2281 GGTYTFRTACVSKAGMGYSSPSEQVLLGGPSHLASEEESQGRSAQPLPSTKTAFOTQI 2340
QY 7679 QRGFSVVRQWEKASGRALAAKIIPYHPKDKTAVLREYKALKGLRPHLAQLHAAVLSP 7738
DB 2341 QRGFSVVRQWEKASGRALAAKIIPYHPKDKTAVLREYKALKGLRPHLAQLHAAVLSP 2400
QY 7739 RHLVLILELCSGPPELLCLAERASYSESEVKDYLMQMSATQYLLHNQHILHLDLRSENMI 7798
DB 2401 RHLVLILELCSGPPELLCLAERASYSESEVKDYLMQMSATQYLLHNQHILHLDLRSENMI 2460
QY 7799 ITEYNLLKVDLGNQAQSLQSKVLPSPDKFYLETMAPELLEGOGAVPQTDIWAIGYTA 7858
DB 2461 ITEYNLLKVDLGNQAQSLQSKVLPSPDKFYLETMAPELLEGOGAVPQTDIWAIGYTA 2520
QY 7859 IMLSAEYPVSBEGADLQRLKGLVRLSRCYAGLSGGAVAFRLSTLCAQPWGRPCASSC 7918

DB 2521 IMLSAEYPVSBEGADLQRLKGLVRLSRCYAGLSGGAVAFRLSTLCAQPWGRPCASSC 2580
QY 7919 LQCPWLTERGPGACSRPAPVTFPTARLVFVRNREKERALLYKRNHNLAAQVR 7968
DB 2581 LQCPWLTERGPGACSRPAPVTFPTARLVFVRNREKERALLYKRNHNLAAQVR 2630
RESULT 5
US-10-307-019--6
; Sequence 6, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenlin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SCIOS, INC.
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DV1
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 2596
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-307-019--6
Query Match 32.8%; Score 13528; DB 14; Length 2596;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 5373 MLERFTPKVKYKGGSSITFSVKVEGRVPVTHMLREAEAGVLWIGPDTPGYTVASSAQOH 5432
DB 1 MLERFTPKVKYKGGSSITFSVKVEGRVPVTHMLREAEAGVLWIGPDTPGYTVASSAQOH 60
QY 5433 SLVLLDVGHQHQGTTCYCIASNAAGQALCSASLHVSGLPKVEQEKKYKEALISTFLOGTTQ 5492
DB 61 SLVLLDVGHQHQGTTCYCIASNAAGQALCSASLHVSGLPKVEQEKKYKEALISTFLOGTTQ 120
QY 5493 AISAQGLETA SFADLGGQKKEPPLAAKEALGHLSLAEVGTEEFLOKLTLSQITEMVSAKT 5552
DB 121 AISAQGLETA SFADLGGQKKEPPLAAKEALGHLSLAEVGTEEFLOKLTLSQITEMVSAKT 180
QY 5553 QAKLVPGGSDSDSKTPSASPRHGRSRPSSIOESSSESDGDARCEIFDIYVVTADYL 5612
DB 181 QAKLVPGGSDSDSKTPSASPRHGRSRPSSIOESSSESDGDARCEIFDIYVVTADYL 240
QY 5613 PLGAEQDAITLREGQVVEVLDAHPRLWLVRKPTKSSPSRQGVSPAYLDRRLKLSPEW 5672
DB 241 PLGAEQDAITLREGQVVEVLDAHPRLWLVRKPTKSSPSRQGVSPAYLDRRLKLSPEW 300
QY 5673 GAABEPFPGAEVSEDEYKARLSSVITQELLSSEQA FVEBLOQLQSHHLQHLRCHPVTIA 5732
DB 301 GAABEPFPGAEVSEDEYKARLSSVITQELLSSEQA FVEBLOQLQSHHLQHLRCHPVTIA 360
QY 5733 VAGQKAVIRNVDRDIGRFHSSFLQELQQCDTDDDVAMCFIKQAQAFQYLFVLGVRQAE 5792
DB 361 VAGQKAVIRNVDRDIGRFHSSFLQELQQCDTDDDVAMCFIKQAQAFQYLFVLGVRQAE 420
QY 5793 SVVYSTAIQEFYKKAABEALLAGDPSPQPPPLQHYLEQPVVRVQRYQALLKELRNKAR 5852
DB 421 SVVYSTAIQEFYKKAABEALLAGDPSPQPPPLQHYLEQPVVRVQRYQALLKELRNKAR 480
QY 5853 NRONCALLQCAVAVSALPQRAENKLHVSLMENYPTGLEALGEPTRQGHFIWEGAPGAR 5912
DB 481 NRONCALLQCAVAVSALPQRAENKLHVSLMENYPTGLEALGEPTRQGHFIWEGAPGAR 540
QY 5913 MPWKGNHRVFLFRNHLVICKPRRSRTDTVSYVFRNMMLSSIDLNDQVEGDRAFEVW 5972

Db 541 MPWKGNHNVLFNHLVCKPRDSDTDSVYFRNMWKLSSLDLNDQVGGDRAPFVW 600
 QY 5973 QERDSVRKYLLOARTAIKSSWKEICGQOORLALPVRPDPPEELADCTAELGETVK 6032
 Db 601 QERDSVRKYLLOARTAIKSSWKEICGQOORLALPVRPDPPEELADCTAELGETVK 660
 QY 6033 LACVTCGPKPVIWYKDGKAVODPHHILTEDPDGSCALILDSLTGDSQVWCFRASA 6092
 Db 661 LACVTCGPKPVIWYKDGKAVODPHHILTEDPDGSCALILDSLTGDSQVWCFRASA 720
 QY 6093 AGNSTGLKILVQVPPFRFVNVKVRASPFVEGEDAQTCTIEGAPYQIRWYKDGALLTGN 6152
 Db 721 AGNSTGLKILVQVPPFRFVNVKVRASPFVEGEDAQTCTIEGAPYQIRWYKDGALLTGN 780
 QY 6153 KFOTLSEPRSGLLVIRAAKEDLGYECELVNRLSARASAEALRQSPMLQAQSOCHR 6212
 Db 781 KFOTLSEPRSGLLVIRAAKEDLGYECELVNRLSARASAEALRQSPMLQAQSOCHR 840
 QY 6213 EQLVAAVEDTTLERADQEVTSVLRKLLGPKAPGSTGDLTGPGRGAPALQETGSQPP 6272
 Db 841 EQLVAAVEDTTLERADQEVTSVLRKLLGPKAPGSTGDLTGPGRGAPALQETGSQPP 900
 QY 6273 VTGTSEAPAVPRVPOPLLHEGPEQBPPEA IARAQEWTPVIRMEGAAMPAGAGTCELLWDVH 6332
 Db 901 VTGTSEAPAVPRVPOPLLHEGPEQBPPEA IARAQEWTPVIRMEGAAMPAGAGTCELLWDVH 960
 QY 6333 SHVRETTQRTYTYCAIDTHTARPPSMQVITEDVQAGTGTGAQPEALIEGDPQPSVTWYK 6392
 Db 961 SHVRETTQRTYTYCAIDTHTARPPSMQVITEDVQAGTGTGAQPEALIEGDPQPSVTWYK 1020
 QY 6393 DSQVLQSTRSQOQEGTYSVLIRHVASKDAGVYTCLAQNTGGQVLCKAELIIVLGDNE 6452
 Db 1021 DSQVLQSTRSQOQEGTYSVLIRHVASKDAGVYTCLAQNTGGQVLCKAELIIVLGDNE 1080
 QY 6453 PDSKQSHRRKLSHFYEVKEIGRGVFKVQKGNKILCAAKTIPLSRSTRAQAYRE 6512
 Db 1081 PDSKQSHRRKLSHFYEVKEIGRGVFKVQKGNKILCAAKTIPLSRSTRAQAYRE 1140
 QY 6513 RDILAALSHPLVTLGLDQFETKTLILILELCSSEELLRLYRGVVTEAEVYVYIQQLV 6572
 Db 1141 RDILAALSHPLVTLGLDQFETKTLILILELCSSEELLRLYRGVVTEAEVYVYIQQLV 1200
 QY 6573 EGLHYLSHGVLHLDIKPSNLMVHPAREDIKICDFGFAQNIPTAELOFQSYGSPFVSP 6632
 Db 1201 EGLHYLSHGVLHLDIKPSNLMVHPAREDIKICDFGFAQNIPTAELOFQSYGSPFVSP 1260
 QY 6633 EIIQONPVSEASDIWANGVISLTCSSPPAGSDRATLNLVLEGRVSNSSPMAAHLSE 6692
 Db 1261 EIIQONPVSEASDIWANGVISLTCSSPPAGSDRATLNLVLEGRVSNSSPMAAHLSE 1320
 QY 6693 DAKOFIKATLQAPQAPSAACLSHPWFLKSPABEAHINTYKQLFLLARSRWORSUM 6752
 Db 1321 DAKOFIKATLQAPQAPSAACLSHPWFLKSPABEAHINTYKQLFLLARSRWORSUM 1380
 QY 6753 SYKSILVMSRIPPELLRPPDPSLGVARHLCRDTCGSSSSSSSDNELLAPFAAKSLPPS 6812
 Db 1381 SYKSILVMSRIPPELLRPPDPSLGVARHLCRDTCGSSSSSSSDNELLAPFAAKSLPPS 1440
 QY 6813 PVTHSPLHPRGFLRPSASLPEEABASERSTEAPAPASPEGAGPPAAQOCVPRHSVIRS 6872
 Db 1441 PVTHSPLHPRGFLRPSASLPEEABASERSTEAPAPASPEGAGPPAAQOCVPRHSVIRS 1500
 QY 6873 LFYHQAGSPHGGALAPGSRHPRARRHLLKGGVIAGALPGLREPLMEHRVLEEAAREE 6932
 Db 1501 LFYHQAGSPHGGALAPGSRHPRARRHLLKGGVIAGALPGLREPLMEHRVLEEAAREE 1560
 QY 6933 QATLLAKAPSPETALRPAASGTHLAPGSHSLEHDSPTPRPSEACGEAQRPSAPSGG 6992
 Db 1561 QATLLAKAPSPETALRPAASGTHLAPGSHSLEHDSPTPRPSEACGEAQRPSAPSGG 1620
 QY 6993 APIRDMGHVPGQSKQLPSTGGHPGTAQPERSPDPSWGPQAPFCHPKQGSAPQSGCSPHPA 7052

Db 1621 APIRDMGHVPGQSKQLPSTGGHPGTAQPERSPDPSWGPQAPFCHPKQGSAPQSGCSPHPA 1680
 QY 7053 VAPCPGSGFPDGSCKEABLVPSPPFLGQOQAPPAPAKASPLDSDXWGGPDLSLGRPKPG 7112
 Db 1681 VAPCPGSGFPDGSCKEABLVPSPPFLGQOQAPPAPAKASPLDSDXWGGPDLSLGRPKPG 1740
 QY 7113 PCSSPGSAQSSQSSVSLRVGSSQVTEPGPSLDAEGWTOEAEEDLSSTFTLQRPQEA 7172
 Db 1741 PCSSPGSAQSSQSSVSLRVGSSQVTEPGPSLDAEGWTOEAEEDLSSTFTLQRPQEA 1800
 QY 7173 TNRKFSLGRGGYAGVAGYGTFAFGDAGMGLGGQPMWARIAWAVSSEEEQEEAARAS 7232
 Db 1801 TNRKFSLGRGGYAGVAGYGTFAFGDAGMGLGGQPMWARIAWAVSSEEEQEEAARAS 1860
 QY 7233 QSEEEQEEAARASPLQVUSARPVPEVGRAPTRSSPEPTPWEDIGOVSLVQIRDLSDGDAEA 7292
 Db 1861 QSEEEQEEAARASPLQVUSARPVPEVGRAPTRSSPEPTPWEDIGOVSLVQIRDLSDGDAEA 1920
 QY 7293 DTISLDSIENVDPAYLNLSDLYDIKYLPPEFMI FRKVPKSAQPEPPSPMAESELAEFEPT 7352
 Db 1921 DTISLDSIENVDPAYLNLSDLYDIKYLPPEFMI FRKVPKSAQPEPPSPMAESELAEFEPT 1980
 QY 7353 WNPCELGPHAGLEITTESEEDVDALLAAAVGRKRKSSPSRSLFHPGRHLPLDEPAEL 7412
 Db 1981 WNPCELGPHAGLEITTESEEDVDALLAAAVGRKRKSSPSRSLFHPGRHLPLDEPAEL 2040
 QY 7413 GLRERVKASVEHISRLKGRPEGLEKEGPPRKPGLASFRLSGLKSWDRAPTFLRELSD 7472
 Db 2041 GLRERVKASVEHISRLKGRPEGLEKEGPPRKPGLASFRLSGLKSWDRAPTFLRELSD 2100
 QY 7473 TVVLQCSVTILACQVSAQAPAAQATWSKCAPLESSRVLISATLKNFOLLTLVVVAEDIG 7532
 Db 2101 TVVLQCSVTILACQVSAQAPAAQATWSKCAPLESSRVLISATLKNFOLLTLVVVAEDIG 2160
 QY 7533 VYTCVSVALGTVTTTGVLRKAERPSSPCPDIGEVADGVLVWKPVEYGPVTYIVQC 7592
 Db 2161 VYTCVSVALGTVTTTGVLRKAERPSSPCPDIGEVADGVLVWKPVEYGPVTYIVQC 2220
 QY 7593 SLEGGSWTTLASDIPDCCYLTSKLSRGTYTFRACVSKAGMGYPSSPSEQVLLGGPSHL 7652
 Db 2221 SLEGGSWTTLASDIPDCCYLTSKLSRGTYTFRACVSKAGMGYPSSPSEQVLLGGPSHL 2280
 QY 7653 ASEESQGRSQAPLSTTKTFAFQTOIQRGFSVVRQCEKASGRALAAKIIPYHPKDKTA 7712
 Db 2281 ASEESQGRSQAPLSTTKTFAFQTOIQRGFSVVRQCEKASGRALAAKIIPYHPKDKTA 2340
 QY 7713 VLREYALKGRHPLHAQHAAYLSPRHLVILELCSGPELLPCLAEASYSSEVSKDYL 7772
 Db 2341 VLREYALKGRHPLHAQHAAYLSPRHLVILELCSGPELLPCLAEASYSSEVSKDYL 2400
 QY 7773 WQMLSATQYLNQHTLHDLRSENMITEYNLLKVVLDLGNASLSQEKVLPDSKFKDYLE 7832
 Db 2401 WQMLSATQYLNQHTLHDLRSENMITEYNLLKVVLDLGNASLSQEKVLPDSKFKDYLE 2460
 QY 7833 TWAPELLGSGQAVPOTDIWAGVTAFLMSAEYPVSSSEGARDLQGRKGLVRLSRVRE 7892
 Db 2461 TWAPELLGSGQAVPOTDIWAGVTAFLMSAEYPVSSSEGARDLQGRKGLVRLSRVRE 2520
 QY 7893 LSGGAVAFRLSTLCAQWGRCASSCLOCPLWTEGPRACSPAPVPTPTARLRFVVRRE 7952
 Db 2521 LSGGAVAFRLSTLCAQWGRCASSCLOCPLWTEGPRACSPAPVPTPTARLRFVVRRE 2580
 QY 7953 KERALLYKXRNLAQVR 7968
 Db 2581 KERALLYKXRNLAQVR 2596

RESULT 6

US-10-093-463-72
 ; Sequence 72, Application US/10093463
 ; Publication No. US20030208039A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Padigaru, Muralidhara

```

; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glenda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zernusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypeptide
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72

; LENGTH: 4691
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-093-463-72

Query Match      25.5%; Score 10519.5; DB 15; Length 4691;
Best Local Similarity 51.1%; Pred. No. 0;
Matches 2367; Conservative 118; Mismatches 342; Indels 1807; Gaps 66;

QY      2595  MPLYNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTASARLEVRVKKVWFLKALD 2654
Db      1  MPLYNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTASARLEVRVKKVWFLKALD 60

QY      2655  DLSAERGTLALQCEVSDPEAHVVRKQGVQGPSKDYDPLHTAGTRGLVVDVSPEDAG 2714
Db      61  DLSAERGTLALQCEVSDPEAHVVRKQGVQGPSKDYDPLHTAGTRGLVVDVSPEDAG 120

QY      2715  LYTVHVGSEETRVRVVDLHVGITKRLKTMEVLGEGSCSFECVLSHESASDPAMWTVGG 2774
Db      121  LYTVHVGSEETRVRVVDLHVGITKRLKTMEVLGEGSCSFECVLSHESASDPAMWTVGG 180

QY      2775  KTVGSSRFQATRQGRKYILVVRAPSDAGEVWFSVRLGTSKASLIVVEERPAAIKPLE 2834
Db      181  KTVGSSRFQATRQGRKYILVVRAPSDAGEVWFSVRLGTSKASLIVVEERPAAIKPLE 240

QY      2835  DQWVAPGEDVELRCELSRAGTPVHMLKORKATRKQKDYDVCCEGTAMMLVIRGASLKDAG 2894
Db      241  DQWVAPGEDVELRCELSRAGTPVHMLKORKATRKQKDYDVCCEGTAMMLVIRGASLKDAG 300

QY      2895  EYTCEVEASKSTASLHVSEKANCFTBELTNLOVEKGTAVFTCKTEHPAATVTRKGLLE 2954
Db      301  EYTCEVEASKSTASLHVSEKANCFTBELTNLOVEKGTAVFTCKTEHPAATVTRKGLLE 360

QY      2955  LRASGHQPSQGLTLRLTISALEKADSDTYTCDTGOAQSRQAQLLVQGRVHIIELEDV 3014
Db      361  LRASGHQPSQGLTLRLTISALEKADSDTYTCDTGOAQSRQAQLLVQGRVHIIELEDV 420

QY      3015  DVQEGSSATFRCRISPANYEPVHFLDKTPLHANELNEIDAOPGGVHVLTQLQALKDSSG 3074
Db      421  DVQEGSSATFRCRISPANYEPVHFLDKTPLHANELNEIDAOPGGVHVLTQLQALKDSSG 480

QY      3075  TIYFEAGDQASAAALRVTEKPSVFSRELTDAITTEGEDTLVCESTCTDIPMCWTKDGT 3134
Db      481  TIYFEAGDQASAAALRVTEKPSVFSRELTDAITTEGEDTLVCESTCTDIPMCWTKDGT 540

QY      3135  LRGSARCOLSHGHRAQLLITGATLQDSGRYKCEAGGACSSIVRVHARVPVFOEALKDL 3194
Db      541  LRGSARCOLSHGHRAQLLITGATLQDSGRYKCEAGGACSSIVRVHARVPVFOEALKDL 600

QY      3195  EVLEGGAATLRCVLSVAAPVKWYGNVLRPGDKYSLRQEGAMLELVVRNLRPQDSGRY 3254
Db      601  EVLEGGAATLRCVLSVAAPVKWYGNVLRPGDKYSLRQEGAMLELVVRNLRPQDSGRY 660

QY      3255  SCSFGDQTTSATLTVTALPAQFIGKLRNKEATGATATLRCBSKTAPVWNRKGSSTLRD 3314
Db      661  SCSFGDQTTSATLTVTALPAQFIGKLRNKEATGATATLRCBSKTAPVWNRKGSSTLRD 720

QY      3315  GDRYCLRQDGAMCELQIRGLAMVDAAEYSCVCGEERTSASLTIRPMPAHFICRLRHQESI 3374
Db      721  GDRYCLRQDGAMCELQIRGLAMVDAAEYSCVCGEERTSASLTIRPMPAHFICRLRHQESI 780

QY      3375  EGATATLRCBSKAAPVWNRKGRESLRDGRHSRLQDGVAVCELOICGLAVADAGEYSCVC 3434
Db      781  EGATATLRCBSKAAPVWNRKGRESLRDGRHSRLQDGVAVCELOICGLAVADAGEYSCVC 840

QY      3435  GEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWCELSKVAPVWNRKGPENLRDGRY 3494
Db      841  GEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWCELSKVAPVWNRKGPENLRDGRY 900

QY      3495  ILRQEGTRCELOICGLAMADAGEYLCVCCQERTSATLTIRALPARPIEDVKNQOAREGAT 3554
Db      901  ILRQEGTRCELOICGLAMADAGEYLCVCCQERTSATLTIRALPARPIEDVKNQOAREGAT 960
```


Db 3057 VOPDDSDWTVTADGSHHALLRLSRAQPHHAGEVTFACRDVAVASARLTVLGLPDPEDAENV 3116
Qy 4533 AHSHTVTLSWAAPMSGGGLCYRVEVEKAGATQWRLCHELVPGPECVWDGLAGETY 4592
Db 3117 ARSHTVTLSWAAPMSGGGLCYRVEVEKAGATQWRLCHELVPGPECVWDGLAGETY 3176
Qy 4593 RFRVAACVPGAGBPVHLPTQVRLAEPPKVPPOPSAPESQVNAAGDVLSLELVVAEAG 4652
Db 3177 RFRVAACVPGAGBPVHLPTQVRL-EPKVPPOPSAPESQVNAAGDVLCLELVVAEAG 3235
Qy 4653 EVIWHKMERIOPGGRFEVVSQGRQMLVINGFTAEQGEYHCGLAGS:CPAAATFOVA 4712
Db 3236 EVIWHKMERIOPGGRFEVVSQGRQMLVINGFTAEQGEYHCGLAGS:CPAAATFOVA 3295
Qy 4713 LSPASVDEAPQSLPPZAAQBGDLHLWEALARKRMSRREPTLDSISELPEEDGSRORLP 4772
Db 3296 LSPASVDEAPQSLPPZAAQBGDLHLWEALARKRMSRREPTLDSISELPEEDGSRORLP 3355
Qy 4773 QEAEEVAPDLSEGYSTADELARTGDADLSHTSSDDESAGTSPSLVTLKXAGRPSTPLA 4832
Db 3356 QEAEEVAPDLSEGYSTADELARTGDADLSHTSSDDESAGTSPSLVTLKXAGRPSTPLA 3415
Qy 4833 SKVGAP-----NAPSVKPOOQOEPL 4852
Db 3416 SKVSPPNLACKERFPTPRAGRSLGLFGVADPAFFGSEARSACTRCAAPPPRESLKEEP- 3474
Qy 4853 AAVRPPLGLDSTKDLG-----DPSMD-----KAAVKIQ-----AANFK 4884
Db 3475 ASCLP--GAMEAVELARKLOBEATCSICLDYFTDPVMTTCGHNFCRACIQLSWEKARGKK 3532
Qy 4885 GYKVRK-----EMKQGGPMFSHTFGTEAQVGDALRECVKASKADVRARWLKOG 4935
Db 3533 GRRKRGSGFPCREKMSQPNLLFN-----RLTKVAEMA----- 3568
Qy 4936 VELTDGRHHHDQLDGDGTCSLLIAGLDRADAGCYTCOVSNK-----FGVTHSACVVVSGS 4991
Db 3569 -----QOH-----POLQKD-----LCQEHHEPLKLFCKQDSPICVVCRE 3604
Qy 4992 ESEAESSGGELDDAFRAARLRHLRFTKSPAEVSDBELFLSADEG----- 5038
Db 3605 SRE-----HRLHRVL-----PASEAVQGYKLEEDMEVYLREQITRTGN 3643
Qy 5039 --PAPEPPADHQ-TYREDEHICIRREAL-----TEARQAVTRFQEMF 5079
Db 3644 LOAREEQLAEMQGVKERRERI VLEFEKMNLYLVEEQRLLQALETEEBEATSLRESV 3703
Qy 5080 ATL---GTGVEIKLVE-----QGPRR-----VEMCISK-- 5104
Db 3704 ACLDRQHSLELLLLQLLEERTQGLQMLQDMKEPLSRAALLVLIHGMNLVEFPVVSLLP 3763
Qy 5105 -----ETPAVVPPEPLPSLLTSDAAP-----VFLTE 5131
Db 3764 SPLVLIATKAHTQLGPGTFTDPCEPTPLRISPP-PRPS--TEDVFPDATSAYPYLLLYE 3820
Qy 5132 LQNGEVQDGYP-----VSFQCVVTQPMPSVRWPKDGLLEDDHMYNEDQOG 5180
Db 3821 SRQRYLSSSEGGSCFKDRFVAPCAV-QQTA-----FSSGR-----HYWVGXNIT 3868
Qy 5181 GHQLIITAVPADMG-----VYRCIAENSMGVSTKAELRVDLTSTDYDTAADATESSYF 5236
Db 3869 GDALWALGVCRDNVSRKDRPKC-PENGFWV-----VQLSK-----GTYKLSLF 3911
Qy 5237 SAQGYLSRE---QEGTESTDEGLPOVVEELRDLQVAPGTRLAKFQKVKGYGAPRILY 5293
Db 3912 SALTPVMLEPPSHMGIFLDFEAG-----EVSFYSVSDGSHLHTYSQAT--FPGLQOP 3962
Qy 5294 WFKDQOPLTAGAHIRMTGKILHTLEIISVTRDSGQVAAAYISNAMGAAYASARLLVRGP 5353
Db 3963 FFLGAP-----KSGQWISTVTMAGV-KDLATRGAVVTPALGA----- 4001
Qy 5354 DEPEEKPADSVHQLVPPRMILERFTPKKVKGSSITFSKVEGPRVPTVHMLREAEGRV 5413
Db 4002 -----YAPSATETQS-----PAP-----MSPPRAPE--- 4021

Qy 5414 LMIQDTPGYTVASSAQOQSLVLLDVGROHQGYTTCIASNAAGQALCASLHVSLPKVE 5473
Db 4022 ----BEHPG--VPSLAPRSA-----RACAAAPGSPSPRAA 4051
Qy 5474 EOEKYEALISTFLOGTTQALISAOGLETASFADLGGORKEEPLAAAKEALGHLSLAEVGT 5533
Db 4052 EAARRPADSTAFPL-SVRMAA-----PDLSTNLQEEATCA- 4087
Qy 5534 EFLQKLTSTIEMVSAKITQAKLOVPGGSDSDSKTSPAS--PRHGRSRPSSSI----- 5585
Db 4088 ICLDYFTDVTWDCGHNFCREICRCWGOPEARTAPSAASCRRGTGCTGPTARLWRPFW 4147
Qy 5586 QSSSESESDGARGEIFDIYVVTADYLPGLAB-QDAITL-----REGQYVEVLDAHPL 5638
Db 4148 RGACTRRRSRR-----VPAHREPLAFCGDELRLCAACERSGEH----- 4189
Qy 5639 RWL-----VRTPTKSSSRQG-----WVSPAYLDRLK-LSPF----- 5671
Db 4190 -WAHRVGRCTRPTKTSRPLEAGTMAANETLLSGAKLEKSLHRLKQMDALLFOAQDET 4248
Qy 5672 ---WGAAEAPPEPGEAVSEDEYKARLSSVIOELLSQAFVEELOFLOSHHL---QHLE 5724
Db 4249 CULWQAEDEGGEORQNVIREFE---RLRLLAEGGTAARAEAGEELKQSAHLAELE 4305
Qy 5725 RCPHVPIAVAGOKA 5738
Db 4306 R--PLPAACAGAAA 4317

RESULT 7

US-10-093-463-74
; Sequence 74, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1 Antibodies that Bind to Antigenic Polypept.
; TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092


```
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 74
; LENGTH: 4675
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-093-463-74

Query Match      25.5%; Score 10506; DB 15; Length 4675;
Best Local Similarity 56.1%; Pred. No. 0;
Matches 2252; Conservative 45; Mismatches 129; Indels 1586; Gaps 35;

QY 2595 MPLYNDSPFHEISHKGRHTLVLSIQRADAGIVRASSLKYSTSARLEVRKPVVFLKALD 2654
Db 1 MPLYNDSPFHEISHKGRHTLVLSIQRADAGIVRASSLKYSTSARLEVRKPVVFLKALD 60

QY 2655 DLSAEEGRGTLALQCEVSDPEAHVVRKDGVLQGPSDKYDFLHTAGTRGLVHVHDSPEADG 2714
Db 61 DLSAEEGRGTLALQCEVSDPEAHVVRKDGVLQGPSDKYDFLHTAGTRGLVHVHDSPEADG 120

QY 2715 LYTCHVGSSETRARVRVHDHLVGTITKLTMEVLGESCFCFVLSHESASDPAMTVGG 2774
Db 121 LYTCHVGSSETRARVRVHDHLVGTITKLTMEVLGESCFCFVLSHESASDPAMTVGG 180

QY 2775 KTVGSSSRFQATROGRKYILVREAPSDAGEVVFVSRGLTSKASLTVRRPAAIIKPLE 2834
Db 181 KTVGSSSRFQATROGRKYILVREAPSDAGEVVFVSRGLTSKASLTVRRPAAIIKPLE 240

QY 2835 DOWVAPGDELCELRSAGTPVHVKDKRAIKRSOKYDVVCEGTWAMLVIRGASLKDAG 2894
Db 241 DOWVAPGDELCELRSAGTPVHVKDKRAIKRSOKYDVVCEGTWAMLVIRGASLKDAG 300

QY 2895 EYTCEVEASKSTASLHVEEKANCFTTELTNLQVEEKGTAFTVCTKEHPAATVTVWRKGLLE 2954
Db 301 EYTCEVEASKSTASLHVEEKANCFTTELTNLQVEEKGTAFTVCTKEHPAATVTVWRKGLLE 360

QY 2955 LRASGKHOPSOEGTLTLTSLALEKADSDTYTCDIGQAQSRALLVQGRVHIIEDLV 3014
Db 361 LRASGKHOPSOEGTLTLTSLALEKADSDTYTCDIGQAQSRALLVQGRVHIIEDLV 420
```

```
QY 3015 DYQEGSSATFRCRISIPANYEPVHWFLDKTPLHANELNEIDIAQPGGYHYLTLRQLALKDSG 3074
Db 421 DYQEGSSATFRCRISIPANYEPVHWFLDKTPLHANELNEIDIAQPGGYHYLTLRQLALKDSG 480

QY 3075 TIYFEAGDORASAAALRVTEKPSVFSRELTDAITTEGEDTLTVCTETSTCDIPMCWTKDKGT 3134
Db 481 TIYFEAGDORASAAALRVTEKPSVFSRELTDAITTEGEDTLTVCTETSTCDIPMCWTKDKGT 540

QY 3135 LRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRVHARVPVRFQALKDL 3194
Db 541 LRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRVHARVPVRFQALKDL 600

QY 3195 EVLEGGAAATLRCVLSVAAPVKWCYGNVLPBGDKYSLROEGAMLELVVRNLRPODSGRY 3254
Db 601 EVLEGGAAATLRCVLSVAAPVKWCYGNVLPBGDKYSLROEGAMLELVVRNLRPODSGRY 660

QY 3255 SCSFGDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAPVVEWKGSETLRD 3314
Db 661 SCSFGDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAPVVEWKGSETLRD 720

QY 3315 GDRYCLRODGAMCELOIRGLAWDAAEYSCVCGERTSASLTIRPMPAHFIRGRHQS 3374
Db 721 GDRYCLRODGAMCELOIRGLAWDAAEYSCVCGERTSASLTIRPMPAHFIRGRHQS 780

QY 3375 EGATATLRCELSKAAPVVEWKGSRSLRQDGRHSRLRQDGAVCELOICGLAVADAGEYSCVC 3434
Db 781 EGATATLRCELSKAAPVVEWKGSRSLRQDGRHSRLRQDGAVCELOICGLAVADAGEYSCVC 840

QY 3435 GEERTSATLTVKALPAKFTEGLRNEEAEGATAMLCWELSKVAPVVEWKGSRSLRQDGRY 3494
Db 841 GEERTSATLTVKALPAKFTEGLRNEEAEGATAMLCWELSKVAPVVEWKGSRSLRQDGRY 900

QY 3495 ILRQEGTRCELOICGLANADAGEYLCVCGQERTSATLTIRALPARFIEDVKNQBARSGAT 3554
Db 901 ILRQEGTRCELOICGLANADAGEYLCVCGQERTSATLTIRALPARFIEDVKNQBARSGAT 960

QY 3555 AVLQCELNSAAPVVEWKGSETLRDGRYSRLRQDTQKCELOIRGLAMADTGYSCVCGQER 3614
Db 961 AVLQCELNSAAPVVEWKGSETLRDGRYSRLRQDTQKCELOIRGLAMADTGYSCVCGQER 1020

QY 3615 TSAMLTUTRALPTKFTTEGLRNEEAEGATAVLRCELKSWAPVVEWKGSETLRDGRHSIRQ 3674
Db 1021 TSAMLTUTRALPTKFTTEGLRNEEAEGATAVLRCELKSWAPVVEWKGSETLRDGRHSIRQ 1080

QY 3675 DGARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPKSFIEGLNEEATEGDTATLW 3734
Db 1081 DGARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPKSFIEGLNEEATEGDTATLW 1140

QY 3735 CELSAAAPVVEWKGSETLRDGRHSRLRQDGRSCELOIRGLAVDAGEYSCVCGQERTSAT 3794
Db 1141 CELSAAAPVVEWKGSETLRDGRHSRLRQDGRSCELOIRGLAVDAGEYSCVCGQERTSAT 1200

QY 3795 LTVRALPARFIEDVKNQBARSGATAVLQCELSKAAPVVEWKGSETLRGGRYSRLRQDGR 3854
Db 1201 LTVRALPARFIEDVKNQBARSGATAVLQCELSKAAPVVEWKGSETLRGGRYSRLRQDGR 1260

QY 3855 CELQIHGLSVADTGYSCVCGQERTSATLTVR----- 3886
Db 1261 CELQIHGLSVADTGYSCVCGQERTSATLTVRALPARFTQDLKTEASEGATATLQCELS 1320

QY 3887 ----- 3886
Db 1321 KVAPVVEWKGSETLRDGRYSRLRQDGRCELOIHDLVSADAGEYSCVCGQERTSATLTVR 1380

QY 3887 ----- 3886
Db 1381 ALPARFTEGLRNEEAEGATATLQCELSKAAPVVEWKGSETLRDGRYSRLRQDGRY 1440

QY 3887 ----- 3886
Db 1441 IHGLAMADNGYSCVCGQERTSATLTVRALPARFIEDVKNQBARSGATVTVLQCKLRKAA 1500

QY 3887 ----- 3886
```

| | | | | | | | |
|----|------|---|------|----|------|--|------|
| Db | 1501 | VEVRKGPNTLKDGRYSLKQDTSCELIQIRGLVLIADAGEYSCICEQERTSATILTVRALPA | 1560 | Db | 2550 | -----CGQVNRDAAQAGATRELLIHQLEAKDTGEYACVTGGQKTAASLRVTEPEVTI | 2600 |
| Qy | 3887 | ----- | 3886 | Qy | 4162 | VRGLVDAEVTADDEVEFCEVSRAGATGVQWCLQGLPLOSNEVTEVAVRDGRHITLRLKGG | 4221 |
| Db | 1561 | RFIEDVRNHEATEGATAVLQCELSKAAPVWRKGSSETLRDGRYSLRQDGRCELQIRGL | 1620 | Db | 2601 | VAGLVDAEVTADDEVEFCEVSRAGATGVQWCLQGLPLOSNEVTEVAVRDGRHITLRLKGG | 2660 |
| Qy | 3887 | ----- | 3886 | Qy | 4222 | VTPEADAGTVSFHLGHNASQAQLTVRAPEVTILEPFDVQLSEGQDASFCQRLSRASGOBA | 4281 |
| Db | 1621 | AVEDTGEVLCVCGQERTSATILTVRALPARFIDNMNINQEARREGATATLHCELSKVAPVEWR | 1680 | Db | 2661 | VTPEADAGTVSFHLGHNASQAQLTVRAPEVTILEPFDVQLSEGQDASFCQRLSRASGOBA | 2701 |
| Qy | 3887 | ----- | 3886 | Qy | 4282 | RNALGVPLQANEMNDITVEQGTLLHLTKVTTLEDAGTVSFHVGTCSSSEAQLKVT---- | 4337 |
| Db | 1681 | KGPETLRDGRHSRLQDGRSCELQIRGLAVVDAGEYSCVCGQERTSATILTVRALPARFIE | 1740 | Db | 2702 | -----GVPLQANEMNDITVEQGTLLHLTKVTTLEDAGTVSFHVGTCSSSEAQLKVTAVP | 2756 |
| Qy | 3887 | ----- | 3886 | Qy | 4338 | ----- | 4337 |
| Db | 1741 | DVKQEARREGATAVLQCELSKAAPVWRKGSSETLRGGDRYSLRQDGRCELQIHGLSVAD | 1800 | Db | 2757 | CLVRGLQNVDFPAGEVATFSCEDGPQSAIAVRDGIHFHSLMLSGLVADSGTVIFRAGPLV | 2816 |
| Qy | 3887 | ----- | 3886 | Qy | 4338 | ----- | 4337 |
| Db | 1801 | TGEYSCVCGQERTSATILTVRALPARFTQDLKTKEASEGATATLQCELSKVAPVEMKXGPE | 1860 | Db | 2817 | STAKLLIKDPVVEVVSAMQDLAVEEGSSAELLCCQYSRFPVQATKMDEREVHTDGRHVIIE | 2876 |
| Qy | 3887 | ----- | 3886 | Qy | 4338 | -----AKNTVVRGLNVLEALLEGGEALFE | 4360 |
| Db | 1861 | TLRDGRYSLKQDGRCELQIHDLISVADAGEYSCMCGQERTSATILTVRDCHTLHVMHPYP | 1920 | Db | 2877 | QDMNVARLTFRDPALPCDSGIYSCAAAGTRVVALLOVQAQNTVVRGLNVLEALLEGGEALFE | 2936 |
| Qy | 3887 | ----- | 3886 | Qy | 4361 | COLSOPEVAHAHTWLLDDEPVRTSENAEVVFFENGLRHLLLLKNLRPQDSCRVTFLAGDMV | 4420 |
| Db | 1921 | FQIPGLLKEPETLIYIQIPSPVILFTBGLRNEAMEGATATLQCELSKAAPVEMWKGLE | 1980 | Db | 2937 | CQLSQPEVAHAHTWLLDDEPVRTSENAEVVFFENGLRHLLLLKNLRPQDSCRVTFLAGDMV | 2996 |
| Qy | 3887 | ----- | 3886 | Qy | 4421 | TSAPFLTVR-----GWELEILEPLKNAAVRAGQAQRTCTLSEAVPVGEASWYINGAA | 4472 |
| Db | 1981 | ALRDGKYSLRQDGAVALCELOIHLAMADNGVYSLPARFIEDMRNQKATEGATVTLQCKL | 2040 | Db | 2997 | TSAPFLTVR-----GWELEILEPLKNAAVRAGQAQRTCTLSEAVPVGEASWYINGAA | 3056 |
| Qy | 3887 | ----- | 3886 | Qy | 4473 | VQDDSDMTVADGSHQALLLSAQPHHAGEVTFACRDVAVASARLTVLGIPEPPEAEVY | 4532 |
| Db | 2041 | RKAAPVEMWKGPNLTKDGRYSLKQDGTSCELQIRGLVIADAGEYSCICEQERTSATILTV | 2100 | Db | 3057 | VQDDSDMTVADGSHQALLLSAQPHHAGEVTFACRDVAVASARLTVLGIPEPPEAEVY | 3116 |
| Qy | 3887 | ----- | 3886 | Qy | 4533 | AHSSHTVTLISWAAPMDSGGGLCGYRVEKVGATGQWRLCHELVPGPECVVDGLAPGETY | 4592 |
| Db | 2101 | RALPARFIEDVRNHEATEGATAVLQCELSKAAPVEMWKGSETLRDGRYSLRQDGRCEL | 2160 | Db | 3117 | ARSHHTVTLISWAAPMDSGGGLCGYRVEKVGATGQWRLCHELVPGPECVVDGLAPGETY | 3176 |
| Qy | 3887 | ----- | 3886 | Qy | 4593 | RFRVAAVGPVAGEPVHLPQTVRLAPPPQPSAPSPQPSAPSRQVAAGEDVSLLEVVAAEAG | 4652 |
| Db | 2161 | QIRGLAVEDTGEYLCVCGQERTSATILTVRALPARFIDNMNINQEARREGATATLHCELSKVA | 2220 | Db | 3177 | RFRVAAVGPVAGEPVHLPQTVRLAPPPQPSAPSPQPSAPSRQVAAGEDVSLLEVVAAEAG | 3235 |
| Qy | 3887 | ----- | 3886 | Qy | 4653 | EVTHKGMERIQPGRFEVVSQGRQQLVIGKFTAEQDQGEYHCGLAQGSICFAAATFOVA | 4712 |
| Db | 2221 | PVEMWKGPNLTKDGRYSLKQDGTSCELQIRGLSVADAGEYSCVCGQERTSATILTVRALP | 2280 | Db | 3236 | EVTHKGMERIQPGRFEVVSQGRQQLVIGKFTAEQDQGEYHCGLAQGSICFAAATFOVA | 3295 |
| Qy | 3887 | ----- | 3886 | Qy | 4713 | LSPASVDEAPQPSLPPEAAQEGDLHLLWEALARKRMSREPTLDSISELPEEDGRSQRLP | 4772 |
| Db | 2281 | AKFTKGLRNEATEGATAMLQCELSKVAPVWRKGPETLRDGRYNLRQDGRCELQIHG | 2340 | Db | 3296 | LSPASVDEAPQPSLPPEAAQEGDLHLLWEALARKRMSREPTLDSISELPEEDGRSQRLP | 3355 |
| Qy | 3887 | ----- | 3886 | Qy | 4773 | QBAEVAAPDISEGYSTADELARTGDLADLSSDSSDRAGTSLVTVLKKAGPPTSPLA | 4832 |
| Db | 2341 | LSVADTGEYSCVCGQERTSATILTVKAPQVFPFEPQLSLQAEBSGSTATLQCELSSEPTATV | 2400 | Db | 3356 | QBAEVAAPDISEGYSTADELARTGDLADLSSDSSDRAGTSLVTVLKKAGPPTSPLA | 3415 |
| Qy | 3922 | WSKGGIQLQANGRRPRLQCTAELVLQDQREDTGEYTCGSOATSATILTVTAAPVRF | 3981 | Qy | 4833 | SKVGAP-----AAPSVPKQOOQOPEL | 4852 |
| Db | 2401 | WSKGGIQLQANGRRPRLQCTAELVLQDQREDTGEYTCGSOATSATILTVTAAPVRF | 2460 | Db | 3416 | SKVSPNKLACKERFPTPRAGRSLLGTVGADPAFFGERSARCTRRCAAPPPRESLKRREP | 3474 |
| Qy | 3982 | LRELOHQEVEDEGTAHLCCELSRAGASVEMWKGSLQFPCKAYQMVQDGAALLLVRGVE | 4041 | Qy | 4853 | AAVRPPIGLDITKDLG-----DPSMD-----KAAVKIQ-----AAFK | 4884 |
| Db | 2461 | LRELOHQEVEDEGTAHLCCELSRAGASVEMWKGSLQFPCKAYQMVQDGAALLLVRGVE | 2520 | Db | 3475 | ASCLIP--GAMEAVEALARKLOEATCSICLDYFTDPVMTTCGHNFRCACIQLSWEKARGKK | 3532 |
| Qy | 4042 | QEDAGDYTCDTGHTQSMASLSVRPFPKFKRLQSLQEQETGDIARLCCOLSDAESGAVVQ | 4101 | Qy | 4885 | GKVRK-----ENKQOEGPMFSHTFGDTEAQVGDALRLSCVVASKADVARMLKDG | 4935 |
| Db | 2521 | QEDAGDYTCDTGHTQSMASLSVRGGR-----GAA-- | 2549 | Db | 3533 | GRKRKSGFPCPECREMSPORNLLPN-----RLITKVAEVA----- | 3568 |
| Qy | 4102 | WLKEGVELHAGPYEVRSGQATRELLIHQLEAKDTGEYACVTGGQKTAASLRVTEPEVTI | 4161 | Qy | 4936 | VELTDGRHHHIDQLGDGTCSLIAGLDRADAGCYTCQVSNK-----FQVTHSACVVVSGS | 4991 |
| | | | | Db | 3569 | -----QQH-----PGLQKQD-----LCQEHHEPLKLFQCKQDQSPICVVCRE | 3604 |

| | | | |
|----|------|---|------|
| Qy | 4992 | ESEAESSGGELDDAFRAARRHLRFTKSPAESDEELFSADEG----- | 5038 |
| Db | 3605 | SRE-----HLSRV-----PAEAVGGYKLUEDMEYLRQIITGN | 3643 |
| Qy | 5039 | --PAPEEPADWQ--TYREDEHFCIRFEAL-----TEARQAVTRFQEMF | 5079 |
| Db | 3644 | LQAREOSLAEWQGVKKERRIVLEFEKNMLXVLEEEOBRLLOALETEBETASRLRESV | 3703 |
| Qy | 5080 | ATL---GIGVEIKLVE---QOPPR-----VEMCTSK-- | 5104 |
| Db | 3704 | ACLDRQHSLELLLLQLEERSTQGLQMDKMEPLSRAALLVYLHGNNLVFFPVVSLP | 3763 |
| Qy | 5105 | -----ETPAVPVPEPLPSLLTSDAAP-----VFLTE | 5131 |
| Db | 3764 | SPLYLIATKAHTQLGPGTPTFDECPTEPLFISPP-PPS--TEDVVPDATSAVPYLLLYE | 3820 |
| Qy | 5132 | LQHQEVODGVP-----VSFDCVVTGQPMPSVRWFKDGLKEEDDHY | 5172 |
| Db | 3821 | SRGRYLGSPFEGSGFGSKDRFYVAPCAV-GQTA-----FSSGR-----HY | 3860 |

RESULT 8

[illegible]

Db 901 PAATVTWRKGLLELRASGKQPSQEGJTLTLTSALEKADSDTYTCDIGQAQRAQLLVQ 960
Qy 3002 GRRVHIIEDEDVDVQSGSATFRCRISIPANYEPVHWFLLDKTPLHANELNEIDAQPGGYH 3061
Db 961 GRRVHIIEDEDVDVQSGSATFRCRISIPANYEPVHWFLLDKTPLHANELNEIDAQPGGYH 1020
Qy 3062 VLTURQALXDSGTIYFAGDQASAAALRVTEKPSVPSRELTATITGEGEDTLVCTETST 3121
Db 1021 VLTURQALXDSGTIYFAGDQASAAALRVTEKPSVPSRELTATITGEGEDTLVCTETST 1080
Qy 3122 CDIPWCTKDGKTLRGSGARCOLSHEGHRALLITGATLQDSGRYKCEAGACSSSIVRVH 3181
Db 1081 CDIPWCTKDGKTLRGSGARCOLSHEGHRALLITGATLQDSGRYKCEAGACSSSIVRVH 1140
Qy 3182 ARPVRFOEALKDLEVLBGGATLRCVLSSVAAPVKWCYGNVLRPGDKYSLROEGAMLEL 3241
Db 1141 ARPVRFOEALKDLEVLBGGATLRCVLSSVAAPVKWCYGNVLRPGDKYSLROEGAMLEL 1200
Qy 3242 VYRNLRPQDSGRYSCSGDQTTSATLTVTALPAQFICKLRNKEATGATATLCELSKTA 3301
Db 1201 VYRNLRPQDSGRYSCSGDQTTSATLTVTALPAQFICKLRNKEATGATATLCELSKAA 1260
Qy 3302 PVWRKGSSETLRDGRYCLRODGAMCELQIRGLAMVDAEYSCVCGEERTSASLTIRPMP 3361
Db 1261 PVWRKGSSETLRDGRYCLRODGAMCELQIRGLAMVDAEYSCVCGEERTSASLTIRPMP 1320
Qy 3362 AHTIGRLRHQSISGATATLCELSKAAAPVWRKGRSLRDGRHSIRODGAVCELOICG 3421
Db 1321 AHTIGRLRHQSISGATATLCELSKAAAPVWRKGRSLRDGRHSIRODGAVCELOICG 1380
Qy 3422 LAVADAGYSVCVCGEERTSATLTVKALPAKTEGLRNEEAVEGATAMLCELSVAAPVEW 3481
Db 1381 LAVADAGYSVCVCGEERTSATLTVKALPAKTEGLRNEEAVEGATAMLCELSVAAPVEW 1440
Qy 3482 RKGPENLRDGRYILRODGTCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARFI 3541
Db 1441 RKGPENLRDGRYILRODGTCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARFI 1500
Qy 3542 EDVKNQAEAGATVLOCELSAAPVWRKGSSETLRDGRYSLRDGTCKEQLQIRGLAMA 3601
Db 1501 EDVKNQAEAGATVLOCELSAAPVWRKGSSETLRDGRYSLRDGTCKEQLQIRGLAMA 1560
Qy 3602 DTGEYSVCVCGERTSATLTVPALPIKFTTEGLRNEEATEGATATLCELSKAAAPVWRKKGH 3661
Db 1561 DTGEYSVCVCGERTSATLTVPALPIKFTTEGLRNEEATEGATATLCELSKAAAPVWRKKGH 1620
Qy 3662 ETLRDGRHSLRDQGARCELQIRGLVAEDAGEYLCMCKERTSAMLTVRAMPKSFIEGLR 3721
Db 1621 ETLRDGRHSLRDQGARCELQIRGLVAEDAGEYLCMCKERTSAMLTVRAMPKSFIEGLR 1680
Qy 3722 NEEATEGDTATLWCELSKAAAPVWRKGHETLRDGRHSLRDQGRSCELQIRGLAVVDAGE 3781
Db 1681 NEEATEGDTATLWCELSKAAAPVWRKGHETLRDGRHSLRDQGRSCELQIRGLAVVDAGE 1740
Qy 3782 YSCVCGQERTSATLTVPALPIEDVKNQAEAGATVLOCELSKAAAPVWRKGSSETLR 3841
Db 1741 YSCVCGQERTSATLTVPALPIEDVKNQAEAGATVLOCELSKAAAPVWRKGSSETLR 1800
Qy 3842 GDRYSLRDGTCELOICGLHSLVADTGEYSCVCGQERTSATLTVPALPIEDVKNQAEAGATVLOCELSKAAAPVWRKGSSETLR 3901
Db 1801 GDRYSLRDGTCELOICGLHSLVADTGEYSCVCGQERTSATLTVPALPIEDVKNQAEAGATVLOCELSKAAAPVWRKGSSETLR 1860
Qy 3902 EBGSTATLCELSSEPTATVNSKGLQLOANGREPRPLOGCTPAELVLODQREDTGTYTC 3961
Db 1861 SEGATATLCELSK-VAPVWRKGPETLRDGRYSLKQDGTCELOICGLHSLVADAGEYSC 1919
Qy 3962 TCGSQTATLVTAAAPVFLRELQHOEVEEGTATLCELSKAGASVWRKGSGLPPEC 4021
Db 1920 MCGQERTSAMLTVRALPAFTTEGLRNEEAMEGATATLCELSKAA-APVWRKGLBALRDG 1978
Qy 4022 AKYQWQDGAABELLVRGVEQEDAGDYTCDTGHTOSMASLSVRPRPKFKTLQSLQEQET 4081

Db 1979 DKYSLRQDGVAVCELOIHGLAMADNGVYSCVCGQERTSATLTVPALPARFIEDMRNQKATE 2038
Qy 4082 GDIAELCOLSDAESCAVQWLKEGVELHAGPKYEMRSGGATRELLIHQLEAKDTGEVAC 4141
Db 2039 GATVTLQCKLRKA--APVWRKGNLTUKODGRYSLKQDGTSCELQIRGLVIADAGEYSC 2095
Qy 4142 VTGGOKTAASLRVTEPEVTIIVRGLVDAEVTADEDEVEFSCEVSRAGATGVQMCLOGLPLOS 4201
Db 2096 ICEQERTSATLTVPALPARFIEDVENHEATEGATAVLOCELSKAAAP--VWRKGSSETLRD 2153
Qy 4202 NVETVAVRDRGRIHTRLRKGVTPEDAGTVSHLGNHASSAQLTVRAPEVTILEPQDVQL 4261
Db 2154 GDRYSLR-QDGTFRCELOIRGLAVEDETYLCVCGQERTSATLTVPALPARFIDNMNQA 2212
Qy 4262 SSGQASFOCLSRASGOEARWALGVPLQANEMNDITVEOQTLHLLTHKVTLEDAGTV 4321
Db 2213 REGATATLCELSKAAPEV--WRKGRSLRDGRHSLR-QDGAUCELOICGLAVADAGEY 2269
Qy 4322 SFHVGTCSSEAQLKV 4336
Db 2270 SCVCGEERTSATLTV 2284

RESULT 9
US-09-858-664A-2
; Sequence 2..Application US/09858664A
; Patent No. US20020072491A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-2

Query Match 20.4%; Score 8423; DB 9; Length 1665;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6356 PPSMQVTIEDVQACTGTAQFEALIRGDPQPSVTWYKDSVQLVDSTRLSQQQSGTTTSLV 6415
Db 53 PPSMQVTIEDVQACTGTAQFEALIRGDPQPSVTWYKDSVQLVDSTRLSQQQSGTTTSLV 112
Qy 6416 LRHVASKDAGVYTCQAQNTGGQVLCXKAEALLVGLGDNPEPSEKSHRRKXLSHFSYEVKEEIG 6475
Db 113 LRHVASKDAGVYTCQAQNTGGQVLCXKAEALLVGLGDNPEPSEKSHRRKXLSHFSYEVKEEIG 172
Qy 6476 RGVGFGVYKRVQKGNKILCAAKFIPLSRTRAQAYBERDILALSHPLVTGLLDQPETRK 6535
Db 173 RGVGFGVYKRVQKGNKILCAAKFIPLSRTRAQAYBERDILALSHPLVTGLLDQPETRK 232
Qy 6536 TLILILELCSSEELLDRLYKGVVTEAEVKVYIQQVLEGLHYLHSHGVHLHDIKPSNLM 6595
Db 233 TLILILELCSSEELLDRLYKGVVTEAEVKVYIQQVLEGLHYLHSHGVHLHDIKPSNLM 292
Qy 6596 VHPAREDIKTCDFGFAQNTIPAELOFSQYSGSPFVSPETIIQONPVSEASDIWANGVTSYL 6655
Db 293 VHPAREDIKTCDFGFAQNTIPAELOFSQYSGSPFVSPETIIQONPVSEASDIWANGVTSYL 352
Qy 6656 SLTCCSPFAGESDRATLNLVLEGVSWSPMAHLSEDAKDFIKATIQAPQAPPSAAQC 6715
Db 353 SLTCCSPFAGESDRATLNLVLEGVSWSPMAHLSEDAKDFIKATIQAPQAPPSAAQC 412

6836 ABAEBSSTAPAPASPEGAGPPAAGCVPKHSVIRSLFYHCAGSPHGHALAPGSRHP 6895
533 ABAEBSSTAPAPASPEGAGPPAAGCVPKHSVIRSLFYHCAGSPHGHALAPGSRHP 592
6896 ARRRHLKGGYTAGALPGLREPLMEHRVLEEAFAEEQATLLAKAPSFETALRLPASGTH 6955
593 ARRRHLKGGYTAGALPGLREPLMEHRVLEEAFAEEQATLLAKAPSFETALRLPASGTH 652
6956 LAPGHSHLEHDSPTFRSSACGAEALPSPAGSGGAPIRDMGHFPGSKOLPSTGGHPG 7015
653 LAPGHSHLEHDSPTFRSSACGAEALPSPAGSGGAPIRDMGHFPGSKOLPSTGGHPG 712
7016 TAQPERSPDPMGQAPAPCPKQGSAPQEGCSHPAVAPCPGSPFPFGSCKEAPLPSS 7075
713 TAQPERSPDPMGQAPAPCPKQGSAPQEGCSHPAVAPCPGSPFPFGSCKEAPLPSS 772
7076 PFLGQQAAPAKASPLDLSKMGCDISLGRPKPGCSGPGSASQSSQSSVSLRVS 7135
773 PFLGQQAAPAKASPLDLSKMGCDISLGRPKPGCSGPGSASQSSQSSVSLRVS 832
7136 SQVGTPEGSLDAEGWTOEAEDLSSTPLQRPQATMRKFSLGGRGYAGVAGYGTFA 7195
833 SQVGTPEGSLDAEGWTOEAEDLSSTPLQRPQATMRKFSLGGRGYAGVAGYGTFA 892
7196 FGDAGCMLGQGMARIAMWAYSQSEEEBOEAREASQSEBOEAREASPLPOVSARVP 7255
893 FGDAGCMLGQGMARIAMWAYSQSEEEBOEAREASQSEBOEAREASPLPOVSARVP 952
7256 EVGRATRSPTPEWEDIGQVSLVQIRLSDGAEAAADTISLIDISEVDPAYLNSLDYDI 7315
953 EVGRATRSPTPEWEDIGQVSLVQIRLSDGAEAAADTISLIDISEVDPAYLNSLDYDI 1012
7316 KILPEFMTFRKPKSAQPEPSPMAEBELAFPEPTWPGELGPHAGLITESEDD 7375
1013 KILPEFMTFRKPKSAQPEPSPMAEBELAFPEPTWPGELGPHAGLITESEDD 1072
7376 ALLAAVAAGKXKWSRSLFHPFGRHLPLDEPAELGLRERVKASVEHISRLKGRPEG 7435
1073 ALLAAVAAGKXKWSRSLFHPFGRHLPLDEPAELGLRERVKASVEHISRLKGRPEG 1132
7436 LEKEGPRKPKGLASRLSGLSWDRAPTFRLSDETVLQGSVTLACQVSAQPAQAT 7495
1133 LEKEGPRKPKGLASRLSGLSWDRAPTFRLSDETVLQGSVTLACQVSAQPAQAT 1192
7496 WSKDGAPELSSSRVLISATLKNFQLTILVVAEDLGVYTCVSNALGVTTTIGVLRKAE 7555
1193 WSKDGAPELSSSRVLISATLKNFQLTILVVAEDLGVYTCVSNALGVTTTIGVLRKAE 1252
7556 RPSPPCPDIEGYADGVLLWKPVESYGPVTYIVQCSLEGGSWTLASDIFDCCYLTSK 7615
1253 RPSPPCPDIEGYADGVLLWKPVESYGPVTYIVQCSLEGGSWTLASDIFDCCYLTSK 1312
7616 LSRGGTYTTRTACVSKAGMPYSSPSEOVLLGCPSHLASEESQGRSAQPLPSTKTAFO 7675
1313 LSRGGTYTTRTACVSKAGMPYSSPSEOVLLGCPSHLASEESQGRSAQPLPSTKTAFO 1372
7676 TQORGRFSVRCWEKASGRALAKIIPYHPKOTAVLREYALKGRHPLHQAALAA 7735
1373 TQORGRFSVRCWEKASGRALAKIIPYHPKOTAVLREYALKGRHPLHQAALAA 1432
7736 LSPRHVLIILELCSGPELLPCIAERASYSSEVKDYLWQMLSATQYLHNQHLHLDRSE 7795
1433 LSPRHVLIILELCSGPELLPCIAERASYSSEVKDYLWQMLSATQYLHNQHLHLDRSE 1492
7796 NMIITRYNLLKVVDLNAQSLSQEKLVPDXTKDYLETMAPELLEGQAVPOTDIWATGV 7855
1493 NMIITRYNLLKVVDLNAQSLSQEKLVPDXTKDYLETMAPELLEGQAVPOTDIWATGV 1552
7856 TAFIMLSAEYPSVSEGDRLQRLKGLVRLSRCYAGLSGGAVAFRLTLCQAPNGRPCA 7915
1553 TAFIMLSAEYPSVSEGDRLQRLKGLVRLSRCYAGLSGGAVAFRLTLCQAPNGRPCA 1612

7916 SSCLOCPMLTEBGPACSRPAPVTFTARLVFVNRREKRALLYKRNHLAQR 7968
1613 SSCLOCPMLTEBGPACSRPAPVTFTARLVFVNRREKRALLYKRNHLAQR 1665

RESULT 11

US-10-415-011-22
; Sequence 22, Application US/10415011
; Publication No. US20040053394A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: CHAMLA, Nandier K.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: XU, Yuming
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: YAO, Monique G.
; APPLICANT: RAMKOMAR, Jayalaxmi
; APPLICANT: DING, Li
; APPLICANT: TANG, Y. Tom
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: NGUYEN, Damiel B.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LU, Yan
; APPLICANT: YUE, Henry
; APPLICANT: BURFORD, Neil
; APPLICANT: BANDMAN, Olga
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: LAL, Preeti G.
; APPLICANT: RECIPON, Shirley A.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: SWARNAKER, Anita
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: KHAN, Farrah A.
; APPLICANT: ISON, Craig H.
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0262 USN
; CURRENT APPLICATION NUMBER: US/10/415,011
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/47728
; PRIOR FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,410
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/244,068
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60/245,708
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/247,672
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/249,565
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,730
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 60/250,807
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040053394A1 7638121CD1
US-10-415-011-22

Query Match 20.4%; Score 8410; DB 12; Length 1665;
Best Local Similarity 99.8%; Pred No. 0;
Matches 1610; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6356 PPSMQVTIEDVQOTGGTAQFAIIEGDPQPSVWYKOSVOLVDSTRLSQQQEGTYSLV 6415
Db 53 PPSMQVTIEDVQOTGGTAQFAIIEGDPQPSVWYKOSVOLVDSTRLSQQQEGTYSLV 112
QY 6416 LRHVASKDAGVYTCIAQNTGGGVLCCKAELLVGGNEPDSEKSHRRKLHSFYEKIEG 6475
Db 113 LRHVASKDAGVYTCIAQNTGGGVLCCKAELLVGGNEPDSEKSHRRKLHSFYEKIEG 172
QY 6476 RGVGFGVQVQKGNKILCAAKFIPLRSRTRAQYRERDILAALSHPLVTGLDQFETR 6535
Db 173 RGVGFGVQVQKGNKILCAAKFIPLRSRTRAQYRERDILAALSHPLVTGLDQFETR 232
QY 6536 TLILILELCSSELDRLRYKGVTEAEVKYIQOLVEGLHSHGVHLHDIKPSNLM 6595
Db 233 TLILILELCSSELDRLRYKGVTEAEVKYIQOLVEGLHSHGVHLHDIKPSNLM 292
QY 6596 VHPAREDIKIDFGFAQNTITPAELQFQSGPEFVSPPIIOQNPVSEASDIWAGVLSYL 6655
Db 293 VHPAREDIKIDFGFAQNTITPAELQFQSGPEFVSPPIIOQNPVSEASDIWAGVLSYL 352
QY 6656 SLTCSPPAGESDRATLNVLEGRVSWSSPMAHLSEDAKDFIKATLQAPQAPSAQC 6715
Db 353 SLTCSPPAGESDRATLNVLEGRVSWSSPMAHLSEDAKDFIKATLQAPQAPSAQC 412
QY 6716 LSHVFWFLSKMAEAEAHFINTKQLFLLARSRWQSLMSYKSLVMSIPILLRGPDPSPS 6775
Db 413 LSHVFWFLSKMAEAEAHFINTKQLFLLARSRWQSLMSYKSLVMSIPILLRGPDPSPS 472
QY 6776 LGVARHLCRDTGGSSSSSSSDNELAPPARAKSLPPSPVTHSPLLHPRGFLRPSASLPEE 6835
Db 473 LGVARHLCRDTGGSSSSSSSDNELAPPARAKSLPPSPVTHSPLLHPRGFLRPSASLPEE 532
QY 6836 AEASERSTEAPASPAGPAPPAQCVPHSVIRSLFYHQAQESPEHAGALPGRRH 6895
Db 533 AEASERSTEAPASPAGPAPPAQCVPHSVIRSLFYHQAQESPEHAGALPGRRH 592
QY 6896 ARRHLLKGGYIAGALPGLREPLMEHRLVBEAEAREEQATLLAKAPSFETALRLPASGTH 6955
Db 593 ARRHLLKGGYIAGALPGLREPLMEHRLVBEAEAREEQATLLAKAPSFETALRLPASGTH 652
QY 6956 LAPGHSLSLEHDSSTPRPSSEACGEAORLPSPSGCAPTRDMGHGPGSKQLSTGCHPG 7015
Db 653 LAPGHSLSLEHDSSTPRPSSEACGEAORLPSPSGCAPTRDMGHGPGSKQLSTGCHPG 712
QY 7016 TAQPERFSPSPWQAPAFCHPKGSAPOEGCSPPHAPVAPCPGSPPPGSCKEAPLVPS 7075
Db 713 TAQPERFSPSPWQAPAFCHPKGSAPOEGCSPPHAPVAPCPGSPPPGSCKEAPLVPS 772
QY 7076 PFLGQOAPAPAKASPLPSKMGPGDISLPGRKPGCPSSPGSASQASSSQVSSLRVGS 7135
Db 773 PFLGQOAPAPAKASPLPSKMGPGDISLPGRKPGCPSSPGSASQASSSQVSSLRVGS 832
QY 7136 SQVTEPGSLDAEGWTOEADLSSTPTLQRPQEQATMRKESLGGGGVAGVAGTGA 7195
Db 833 SQVTEPGSLDAEGWTOEADLSSTPTLQRPQEQATMRKESLGGGGVAGVAGTGA 892
QY 7196 FGGDAGMLCGGPMWARIAWASQSEEEQAEARASQSEEQAEARASPLPQVSARVP 7255
Db 893 FGGDAGMLCGGPMWARIAWASQSEEEQAEARASQSEEQAEARASPLPQVSARVP 952
QY 7256 EVGRAPTRSSPEPTPWEDIGQVSLVOIRLSDGAEADTISLIDISEVDPAYNLNLDLYDI 7315
Db 953 EVGRAPTRSSPEPTPWEDIGQVSLVOIRLSDGAEADTISLIDISEVDPAYNLNLDLYDI 1012
QY 7316 KYLPFFFMIFRKYKPSAQPPSPMAEEBELAEPPTWPGELGPHAGLEITESESDVD 7375
Db 1013 KYLPFFFMIFRKYKPSAQPPSPMAEEBELAEPPTWPGELGPHAGLEITESESDVD 1072
QY 7376 ALLAEAAVGRKRWSSPSLSLHFFPGHPLDPEAPLGLRERVKASVEHISRLKGRREG 7435
Db 1073 ALLAEAAVGRKRWSSPSLSLHFFPGHPLDPEAPLGLRERVKASVEHISRLKGRREG 1132
QY 7436 LEKEGPPRKKPGLASPRLSGLKSWDRAPTFLRELSDETVVLQSVTLACQVSAQPAQAT 7495

Db 1133 LEXEGPPRKKPGLASPRLSGLKSWDRAPTFLRELSDETVVLQSVTLACQVSAQPAQAT 1192
QY 7496 WSKDGAFLPSSSRVLISATLKNFOLLTILVVAEDLVGVYTCVSNALGTVTTTGVLRKAE 7555
Db 1193 WSKDGAFLPSSSRVLISATLKNFOLLTILVVAEDLVGVYTCVSNALGTVTTTGVLRKAE 1252
QY 7556 RPSSSPCPDIEGVYADGVLLVWKPVESYGPVTYVQCSLEGSWTTLASDIFDCCYLRSK 7615
Db 1253 RPSSSPCPDIEGVYADGVLLVWKPVESYGPVTYVQCSLEGSWTTLASDIFDCCYLRSK 1312
QY 7616 LSRGGTYTTRTACVSKAGMPYSSPSEQVLLGGPSHLASEEBSQGRSAQPLSTKTFAFO 7675
Db 1313 LSRGGTYTTRTACVSKAGMPYSSPSEQVLLGGPSHLASEEBSQGRSAQPLSTKTFAFO 1372
QY 7676 TQIORGRFVSVRQCKEKASGRALAAKIIPYHPKDKTAVLREYEALKGLRHPHLAQLHAAY 7735
Db 1373 TQIORGRFVSVRQCKEKASGRALAAKIIPYHPKDKTAVLREYEALKGLRHPHLAQLHAAY 1432
QY 7736 LSPRHLVLILELCSGPPELLPCLAEASYSSESEVKDYLMQMSATQYLHNOHILHDLRSE 7795
Db 1433 LSPRHLVLILELCSGPPELLPCLAEASYSSESEVKDYLMQMSATQYLHNOHILHDLRSE 1492
QY 7796 NMITEYNLLKVVYDIGNAQSLQSEKVLPSDKFKDYLETMAPELLEGQGAVPQTDIWAIGV 7855
Db 1493 NMITEYNLLKVVYDIGNAQSLQSEKVLPSDKFKDYLETMAPELLEGQGAVPQTDIWAIGV 1552
QY 7856 TAFIMLSAEYVSSSEGDRLQRLKGLVRLSRVAGISGGAVAFRLSTLCAQPNRCPA 7915
Db 1553 TAFIMLSAEYVSSSEGDRLQRLKGLVRLSRVAGISGGAVAFRLSTLCAQPNRCPA 1612
QY 7916 SSCLOCPMLTEBGPACSPAPVPTPTARLRFVVRNKRERALLYKRHNLAQVR 7968
Db 1613 SSCLOCPMLTEBGPACSPAPVPTPTARLRFVVRNKRERALLYKRHNLAQVR 1665

RESULT 12

US-10-182-243-46
; Sequence 46, Application US/10182243
; Publication No. US20040048310A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY D.
; APPLICANT: WHYTE, DAVID
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; APPLICANT: MARTINEZ, RICARDO
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES AND PROTEIN KINASE-LIKE
; FILE REFERENCE: 038602/1366
; CURRENT APPLICATION NUMBER: US/10/182,243
; PRIOR FILING DATE: 2003-07-07
; PRIOR APPLICATION NUMBER: PCT/US01/02337
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
; LENGTH: 1618
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-182-243-46

Query Match 20.4%; Score 8407; DB 12; Length 1618;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1612; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 6357 PSMQVTIEDVQOTGGTAQFAIIEGDPQPSVWYKOSVOLVDSTRLSQQQEGTYSLV 6416
Db 1 PSMQVTIEDVQOTGGTAQFAIIEGDPQPSVWYKOSVOLVDSTRLSQQQEGTYSLV 60
QY 6417 RHVASKDAGVYTCIAQNTGGGVLCCKAELLVGGNEPDSEKSHRRKLHSFYEK 6470
Db 61 RHVASKDAGVYTCIAQNTGGGVLCCKAELLVGGNEPDSEKSHRRKLHSFYEK 120


```

Db      241  |||||||AEDIKIDFGPAQNTTAEQFSGYGFVSPETIIQQNVPSEASDIWAMGVSYLSLT 300
QY      6659  |||||||CSSPPAGESDRATLLNVLEGRVSWSPMAAHLSEDAKDFIKATLQRAPOARPSAAQCLSH 6718
Db      301  |||||||CSSPPAGESDRATLLNVLEGRVSWSPMAAHLSEDAKDFIKATLQRAPOARPSAAQCLSH 360
QY      6719  |||||||PWFILKSPMAEAEAFINTKQLKFLARSRWQSLMSYKSIILVMSIPELLRPPSPSLGV 6778
Db      361  |||||||PWFILKSPMAEAEAFINTKQLKFLARSRWQSLMSYKSIILVMSIPELLRPPSPSLGV 420
QY      6779  |||||||AHLCRDTGGSSSSSSSDNELAPARAKSLPPSPVTHSPILLHPRGFLRPSASIPERAEA 6838
Db      421  |||||||AHLCRDTGGSSSSSSSDNELAPARAKSLPPSPVTHSPILLHPRGFLRPSASIPERAEA 480
QY      6839  |||||||SERSTEAPAPASPEGAGPPAAQGCVPVTHSVIRSLFYHQAGESPEHGALAPGSRHRPARR 6898
Db      481  |||||||SERSTEAPAPASPEGAGPPAAQGCVPVTHSVIRSLFYHQAGESPEHGALAPGSRHRPARR 540
QY      6899  |||||||RHLLKGGYTAGALPGLREPLMEHVRVLEBEAAREEQATLLAKAPSFETALRPLASGTHLAP 6958
Db      541  |||||||RHLLKGGYTAGALPGLREPLMEHVRVLEBEAAREEQATLLAKAPSFETALRPLASGTHLAP 600
QY      6959  |||||||GHSLSLEHDSPTSTRPSSACGEAQRPLSPAGSGAPIRDMGHGQSKQLPSTGHPGTAQ 7018
Db      601  |||||||GHSLSLEHDSPTSTRPSSACGEAQRPLSPAGSGAPIRDMGHGQSKQLPSTGHPGTAQ 660
QY      7019  |||||||PERSPDPSWQOPAPFCHPKGSAPOEGCSBHPAVAPCPGSPPGCKEAPLVPSSPFL 7078
Db      661  |||||||PERSPDPSWQOPAPFCHPKGSAPOEGCSBHPAVAPCPGSPPGCKEAPLVPSSPFL 720
QY      7079  |||||||GQOPAPPAPAKASPLDLSKMGPGDISLPGRKPGPCSPGASASQSSQVSSSLRVGSSQV 7138
Db      721  |||||||GQOPAPPAPAKASPLDLSKMGPGDISLPGRKPGPCSPGASASQSSQVSSSLRVGSSQV 780
QY      7139  |||||||GTEPGSLDAGWGTQEAEDLSDSTPTLQROBOQATMRKFSLGGGGVAGVAGTFAFG 7198
Db      781  |||||||GTEPGSLDAGWGTQEAEDLSDSTPTLQROBOQATMRKFSLGGGGVAGVAGTFAFG 840
QY      7199  |||||||DAGMGLGQPMWARIANAVSSEEEQEARAESQSEEQEARAESPLPOVSARVPVEVG 7258
Db      841  |||||||DAGMGLGQPMWARIANAVSSEEEQEARAESQSEEQEARAESPLPOVSARVPVEVG 900
QY      7259  |||||||RAPTRSSPEPTPWEDIGQVSLVQIRDLSDGAEAADTISLDISEVDPAYLNLSDLYIKYL 7318
Db      901  |||||||RAPTRSSPEPTPWEDIGQVSLVQIRDLSDGAEAADTISLDISEVDPAYLNLSDLYIKYL 960
QY      7319  |||||||PFEEWIFRKVPKSAQPPSPMAEELAEPEPTWMPGELGPHAGLEITEESDEDVALL 7378
Db      961  |||||||PFEEWIFRKVPKSAQPPSPMAEELAEPEPTWMPGELGPHAGLEITEESDEDVALL 1020
QY      7379  |||||||AEEAVGRKRKWSPPSRSLFHPGSHLPLDPEPAELGLRERVKASVEHISRLIKGRPEGLEK 7438
Db      1021  |||||||AEEAVGRKRKWSPPSRSLFHPGSHLPLDPEPAELGLRERVKASVEHISRLIKGRPEGLEK 1080
QY      7439  |||||||EGPRKPKGLASPLSGIKSWDRAPTFLRELSDETIVLGOSVTILACQVSAQPAQAATWSK 7498
Db      1081  |||||||EGPRKPKGLASPLSGIKSWDRAPTFLRELSDETIVLGOSVTILACQVSAQPAQAATWSK 1140
QY      7499  |||||||DGAPLESSSRVLIISATLKNFOLLITILVVAEDLGVTYCSVSNALGTTVTGVLKRAERPS 7558
Db      1141  |||||||DGAPLESSSRVLIISATLKNFOLLITILVVAEDLGVTYCSVSNALGTTVTGVLKRAERPS 1200
QY      7559  |||||||SSPCDDIGEVYADGVLLVWKPVSIGYVTIIVQCSLEGGSWTTLASDIFDCCYLTSLKR 7618
Db      1201  |||||||SSPCDDIGEVYADGVLLVWKPVSIGYVTIIVQCSLEGGSWTTLASDIFDCCYLTSLKR 1260
QY      7619  |||||||GGTYTFTTACVSKAGMGPPYSSPEQVLLGGPSPHSLASEEBSQGRSAOPLPSTKTFAFQTOI 7678
Db      1261  |||||||GGTYTFTTACVSKAGMGPPYSSPEQVLLGGPSPHSLASEEBSQGRSAOPLPSTKTFAFQTOI 1320
QY      7679  |||||||QGRFSPVROQWEKASGRALAAKIIPIYHPKDKTAVLREYALKGLRHPHQAQHAAYLSP 7738

```

```

Db      1321  |||||||QGRFSPVROQWEKASGRALAAKIIPIYHPKDKTAVLREYALKGLRHPHQAQHAAYLSP 1380
QY      7739  |||||||RHVLVILELCSGPPELLPCLAEASYSSESVKDYLMQMLSATQYLNQHHTLHLDLSENNI 7798
Db      1381  |||||||RHVLVILELCSGPPELLPCLAEASYSSESVKDYLMQMLSATQYLNQHHTLHLDLSENNI 1440
QY      7799  |||||||ITEYNLLKVVLDLGNAGSLSQEKVLPSPDKFYLETWAPELLEGGQAVPQTDIWAIGVTAF 7858
Db      1441  |||||||ITEYNLLKVVLDLGNAGSLSQEKVLPSPDKFYLETWAPELLEGGQAVPQTDIWAIGVTAF 1500
QY      7859  |||||||IMLSABEYPVSSSGARDLQRLKGLVRLSRVAGLSGGAVAFRLSTLCAQPMGRPCASSC 7918
Db      1501  |||||||IMLSABEYPVSSSGARDLQRLKGLVRLSRVAGLSGGAVAFRLSTLCAQPMGRPCASSC 1560
QY      7919  |||||||LQCPWLTEEGPACSRPAPVTFTPTARLVFVRNREXRALLYKRHNLAQVR 7968
Db      1561  |||||||LQCPWLTEEGPACSRPAPVTFTPTARLVFVRNREXRALLYKRHNLAQVR 1610

```

RESULT 14

```

US-10-307-019-1
; Sequence 1, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng Wenlin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SciOS, INC.
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DV1
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-307-019-1

```

```

Query Match 17.1%; Score 7073; DB 14; Length 1351;
Best Local Similarity 99.9%; Pred. No. 3.6e-316;
Matches 1349; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      6619  |||||||LQFSQYGSPEFVSPETIIQQNVPSEASDIWAMGVSYLSLTCSPPAGESDRATLLNVLEG 6678
Db      2  |||||||VQFSQYGSPEFVSPETIIQQNVPSEASDIWAMGVSYLSLTCSPPAGESDRATLLNVLEG 61
QY      6679  |||||||RVSWSSPMAAHLSEDAKDFIKATLQRAPOARPSAAQCLSHPWFILKSPMAEAEHINTKOL 6738
Db      62  |||||||RVSWSSPMAAHLSEDAKDFIKATLQRAPOARPSAAQCLSHPWFILKSPMAEAEHINTKOL 121
QY      6739  |||||||KFLARSRWQSLMSYKSIILVMSIPELLRPPSPSLGVARHLCRDITGSSSSSSSDN 6798
Db      122  |||||||KFLARSRWQSLMSYKSIILVMSIPELLRPPSPSLGVARHLCRDITGSSSSSSSDN 181
QY      6799  |||||||ELAPPARAKSLPPSPVTHSPILLHPRGFLRPSASIPERAEASERSTEAPAPASPEGAGPP 6858
Db      182  |||||||ELAPPARAKSLPPSPVTHSPILLHPRGFLRPSASIPERAEASERSTEAPAPASPEGAGPP 241
QY      6859  |||||||AAQCVPRHSVIRSLFYHQAGESPEHGALAPGSRHRPARRHLLKGGYIAGALPGLREPL 6918
Db      242  |||||||AAQCVPRHSVIRSLFYHQAGESPEHGALAPGSRHRPARRHLLKGGYIAGALPGLREPL 301
QY      6919  |||||||MEHVRVLEBEAAREEQATLLAKAPSFETALRPLASGTHLAPCHSHLSLEHDSPTSTRPSEA 6978
Db      302  |||||||MEHVRVLEBEAAREEQATLLAKAPSFETALRPLASGTHLAPCHSHLSLEHDSPTSTRPSEA 361
QY      6979  |||||||CGEAQRPLSAPSGGAPIRDMCHPGQSKQLPSTGHPGTAQPERPSPDPSWQOPAPFCHPK 7038

```

Db 362 CGEACRLPSAPSGGAPITRDMGHGPGSKQLPSTGGHGTGTAQPERPSPDPSWGQFAPFCHPK 421
Qy 7039 QGSAPOGSCSHPAVACPPGSRPPGCKEAPLVSSPFLGQCAPAPAKASPPPLDSKM 7098
Db 422 QGSAPOGSCSHPAVACPPGSRPPGCKEAPLVSSPFLGQCAPAPAKASPPPLDSKM 481
Qy 7099 GPGDISLPGRPKPGPCSPGASQASSQSSVSLRVSSQVGTGPGSLDAEGTQEAEDL 7158
Db 482 GPGDISLPGRPKPGPCSPGASQASSQSSVSLRVSSQVGTGPGSLDAEGTQEAEDL 541
Qy 7159 SDSTFTLQRPQOATMRKFSLGGGGGAGVAGYGTFAFGDAGMLGQGPWARIWAWS 7218
Db 542 SDSTFTLQRPQOATMRKFSLGGGGGAGVAGYGTFAFGDAGMLGQGPWARIWAWS 601
Qy 7219 QSEEEQEEAARASQSEEQEAAESPLQVARSVPVEVGRAPTRSSPEPTPEDIQVQS 7278
Db 602 QSEEEQEEAARASQSEEQEAAESPLQVARSVPVEVGRAPTRSSPEPTPEDIQVQS 661
Qy 7279 LVQIRDSGDAAADTISLDSVDPAVLNLSLDYIKYLPFEPFMIIRKPKVKSAPQPPS 7338
Db 662 LVQIRDSGDAAADTISLDSVDPAVLNLSLDYIKYLPFEPFMIIRKPKVKSAPQPPS 721
Qy 7339 PMAEEELAEFPETWPFCELPAGLEITESEDDVALLAAEAAVGRKRWSSPSRSLFH 7398
Db 722 PMAEEELAEFPETWPFCELPAGLEITESEDDVALLAAEAAVGRKRWSSPSRSLFH 781
Qy 7399 FPGRHLPDEPAELGLRERVKASVEHISRILKGRPEGLEKPGPRKXPGJASFSLGLKS 7458
Db 782 FPGRHLPDEPAELGLRERVKASVEHISRILKGRPEGLEKPGPRKXPGJASFSLGLKS 841
Qy 7459 WDRAPTFRLSDETVLQSVTLACQVSAQAQATWSKDGAFLSSSRVLISATLKNF 7518
Db 842 WDRAPTFRLSDETVLQSVTLACQVSAQAQATWSKDGAFLSSSRVLISATLKNF 901
Qy 7519 QLLTLVVAEDLGYVTCVSNALGTVTTTGLRKAERPSSPCPDIGEVVADGVLLVWK 7578
Db 902 QLLTLVVAEDLGYVTCVSNALGTVTTTGLRKAERPSSPCPDIGEVVADGVLLVWK 961
Qy 7579 PVESGPTVYIVQCSLEGGSWTLASDIPDCYILTSKLSRGTYTFRACVSKAGMPYS 7638
Db 962 PVESGPTVYIVQCSLEGGSWTLASDIPDCYILTSKLSRGTYTFRACVSKAGMPYS 1021
Qy 7639 SPSEQVLGGPSHLASEESQGRSAQPLPSYKTFATQIQGRFVSVRQCEKASGRAL 7698
Db 1022 SPSEQVLGGPSHLASEESQGRSAQPLPSYKTFATQIQGRFVSVRQCEKASGRAL 1081
Qy 7699 AAKIIPYHPKXTAVLREYALKGLRHPHLAQLHAAYLSPRHLVLIILELCSGPPELLPCLA 7759
Db 1082 AAKIIPYHPKXTAVLREYALKGLRHPHLAQLHAAYLSPRHLVLIILELCSGPPELLPCLA 1141
Qy 7759 ERASYESSEVKDYLWQMSATQYLHNOHLHLDLSENNMITEYNLLKVVLDLGNASLSQ 7818
Db 1142 ERASYESSEVKDYLWQMSATQYLHNOHLHLDLSENNMITEYNLLKVVLDLGNASLSQ 1201
Qy 7819 EXVLPSDKFKDYLETWAPELLEGGQAVPQTDIWAIGVTAFIMLSAEYPVSEGARDLQRG 7878
Db 1202 EXVLPSDKFKDYLETWAPELLEGGQAVPQTDIWAIGVTAFIMLSAEYPVSEGARDLQRG 1261
Qy 7879 LRKGLVRLSRCVAGLSGGAVAFRLSTLCAQWGRPCASSCLOCPLWITEGACSRPAPVT 7938
Db 1262 LRKGLVRLSRCVAGLSGGAVAFRLSTLCAQWGRPCASSCLOCPLWITEGACSRPAPVT 1321
Qy 7939 FPTARLVRVVRNREKRALLYKRNHNAQVR 7968
Db 1322 FPTARLVRVVRNREKRALLYKRNHNAQVR 1351

RESULT 15
US-10-408-765A-992
; Sequence 992, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.

; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 992
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-408-765A-992

Query Match 16.0%; Score 6619.5; DB 16; Length 1596;
Best Local Similarity 81.7%; Pred. No. 3.1e-295; Indels 21; Gaps 7;
Matches 1309; Conservative 70; Mismatch 203;

Qy 2726 RARVRVHDLHVGTIKELKTMEVLEGESECFVLSHESASDPAMTVGGKTVGSSSRFOA 2785
Db 1 RARVRVHDLHVGTIKELKTMEVLEGESECFVLSHESASDPAMTVGGKTVGSSSRFOA 60
Qy 2786 TROGRKIIVREAAAPSDAGEVVFVSRGLTSKASLIVRRPAAIKPLEDOMVAGEDVE 2845
Db 61 TROGRKIIVREAAAPSDAGEVVFVSRGLTSKASLIVRRPAAIKPLEDOMVAGEDVE 120
Qy 2846 LRCELSRAGTPVHVLKDKRAIKRSQYDVVCGTMAWLVIRGASLKDAGEYTCVEBASKS 2905
Db 121 LRCELSRAGTPVHVLKDKRAIKRSQYDVVCGTMAWLVIRGASLKDAGEYTCVEBASKS 180
Qy 2906 TASLHVEEKANCFTBELTNLQVEEKTAFTCTKTEHPAATVTVRKGLELRSKQHPQ 2965
Db 181 TASLHVEEKANCFTBELTNLQVEEKTAFTCTKTEHPAATVTVRKGLELRSKQHPQ 240
Qy 2966 EGTLSLTLSALEKADSDTYTCDIGQAQRAQLLVQGRVHHIEDLEDVDVQSGSATFR 3025
Db 241 EGTLSLTLSALEKADSDTYTCDIGQAQRAQLLVQGRVHHIEDLEDVDVQSGSATFR 300
Qy 3026 CRISPANYEPVHFLDKTPLHANELNEIDAQPGYHVLTLRQLALKDSGTIYFEAGDQA 3085
Db 301 CRISPANYEPVHFLDKTPLHANELNEIDAQPGYHVLTLRQLALKDSGTIYFEAGDQA 360
Qy 3086 SAALRVTEKPSVSRBELTDATITEGEDTLVCTSTCDIPMCWTGDKTLRGSARCOLSH 3145
Db 361 SAALRVTEKPSVSRBELTDATITEGEDTLVCTSTCDIPMCWTGDKTLRGSARCOLSH 420
Qy 3146 EGHRAQLLITGATLQDSGRYKCBAGACSSSIVRVHARPVRFOEALKDLEVLGGGAATLR 3205
Db 421 EGHRAQLLITGATLQDSGRYKCBAGACSSSIVRVHARPVRFOEALKDLEVLGGGAATLR 480
Qy 3206 CVLSSVAAPVKWCYGNVLPBGDKYSLRQEGAMELIVRNLRPDQSGRYSCSGDQTTSA 3265
Db 481 CVLSSVAAPVKWCYGNVLPBGDKYSLRQEGAMELIVRNLRPDQSGRYSCSGDQTTSA 540
Qy 3266 TLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAPEVWRKGSSETLRDGDYCLRDGA 3325
Db 541 TLTVTALPAQFIGKLRNKEATEGATATLRCELSKTAPEVWRKGSSETLRDGDYCLRDGA 600
Qy 3326 MCELQIRGLAVDAAEYSVCVGBERTSASLTIRMPAHFTIGRLRHQESIEGATATLRCEL 3385
Db 601 MCELQIRGLAVDAAEYSVCVGBERTSASLTIRMPAHFTIGRLRHQESIEGATATLRCEL 660
Qy 3386 SKAAPVWRKGRSLRDGDRHSLRDGAVCELOICGLAVADAGEYSCVCGEERTSATLTV 3445
Db 661 SKAAPVWRKGRSLRDGDRHSLRDGAVCELOICGLAVADAGEYSCVCGEERTSATLTV 720
Qy 3446 KALPAKFTTEGLRNEEAVEGATAMLWCELSKEVAPVWRKGPENLRDGDYTLRQEGTRCEL 3505

Db 721 KALPAKFTGLRNEBAVEGATAMLCELSKVAPVEWRKGPENLRDGRYILRQEGTRCEL 780
QY 3506 QICGLAMADAGEYLCVCGQERTSATLITRALPARFIEDVKQOEAREGATAVLOCELNSAA 3565
Db 781 QICGLAMADAGEYLCVCGQERTSATLITRALPARFIEDVKQOEAREGATAVLOCELNSAA 840
QY 3566 PVWRKGSSTLRDGRYSRLRQDGTQKCELOIRGLAMADTGEYSCVCGQERTSAMLTVRALP 3625
Db 841 PVWRKGSSTLRDGRYSRLRQDGTQKCELOIRGLAMADTGEYSCVCGQERTSAMLTVRALP 900
QY 3626 IKFTTEGLRNEBATEGATVLRCELSKMAPVWKKHETLRDGRHSRLRQDGRARCELOIRG 3685
Db 901 IKFTTEGLRNEBATEGATVLRCELSKMAPVWKKHETLRDGRHSRLRQDGRARCELOIRG 960
QY 3686 LVAEDAGEYLCMCGKERTSAMLTVRAMPKFTIEGLRNEEATEGDTATLWCELSKAAAPVEW 3745
Db 961 LVAEDAGEYLCMCGKERTSAMLTVRAMPKFTIEGLRNEEATEGDTATLWCELSKAAAPVEW 1020
QY 3746 RKGHETLRDGRHSRLRQDGRARCELOIRGLAVVDAGEYSCVCGQERTSATLITVRALPARFI 3805
Db 1021 RKGHETLRDGRHSRLRQDGRARCELOIRGLAVVDAGEYSCVCGQERTSATLITVRALPARFI 1080
QY 3806 EDVKQOEAREGATAVLOCELNSAAAPVEWRKGSSETLRGGDRYSRLRQDGTQKCELOIHGLSVA 3865
Db 1081 EDVKQOEAREGATAVLOCELNSAAAPVEWRKGSSETLRGGDRYSRLRQDGTQKCELOIHGLSVA 1140
QY 3866 DTGEYSCVCGQERTSATLITVRAPQVFPREPLQSOABEGSTATLOCELSBPATATVWVMSKG 3925
Db 1141 DTGEYSCVCGQERTSATLITVRALPARFTQDLKTEASEGATATLOCELSK-VAPVEWVWKG 1199
QY 3926 GLQLQANGRRPRLOGCTAEVLQDLOREDTGEYTCGSOATSATLITVTAAPVRLREL 3985
Db 1200 PETLRDGRYSRLRQDGTQKCELOIHGLSVAADAGEYSCVCGQERTSATLITVRALPARFTEGL 1259
QY 3986 QHVEVDEGGTAHLCELSRAGASVEWRKGSLOLPFCAYQMVDGAAAEILLVRGVEQEDA 4045
Db 1260 RNEEAMEGATATLOCELSKA-APVEWRKGLEALRDGDKYSRLRQDGAVALCELOIHGLAMADN 1318
QY 4046 GDYTCDTGHTOSMASLSVRVPRPKFTLRLOSLQETGDIARLCCOLSDAESGAVVOWLKE 4105
Db 1319 GYVSCVCGQERTSATLITVRALPARFIEDMRNQKATEGATVTLQCKLRKA---APVEWRKG 1375
QY 4106 GVLEHAGPKYEMRSQGAATRELLIHQLEAKDTGEYACVTGGQKTAASLRVTEPEVTIVRGL 4165
Db 1376 PNTLRDGRYSRLRQDGTQKCELOIRGLVADAGEYSCICEQERTSATLITVRALPARFIEDV 1435
QY 4166 VDAEVTADEDEVEFCEVSRAGATGVQNCLOGLPLOSNEVTEVAVRQGRHITLRLKGVTP 4225
Db 1436 RNHEATEGATVLOCELSKAAAP--VEWRKSETLRDGRYSRLR-QDGTQKCELOIRGLAVE 1492
QY 4226 DAGTVSFHLGNHASSAGITVRAPVITILEPLQDVOLSEGOQDASFOCRLSRASGOEARWAL 4285
Db 1493 DTGEYLCVCGQERTSATLITVRALPARFIDNNTQOEAREGATATLHCELSKVAAPVE--WRK 1550
QY 4286 GGVPLQANEMNDITVEQGTLLHLTLHKVTLDEAGTVSFHVGTG 4328
Db 1551 GPETLRDGRHSRLRQEN-----RLNPGGCGCSELGSC 1582

Search completed: September 13, 2004, 12:12:13
Job time : 541 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 14:16:08 ; Search time 1570.02 Seconds
(without alignments)
10628.572 Million cell updates/sec

Title: US-10-077-130-4_COPY_16862_17246

Perfect score: 385
Sequence: 1 ccagggagagatttgaca.....acctgcagcactggagcgc 385

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 347072 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_lm:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|--------------------|
| 1 | 385 | 100.0 | 4901 | 9 | HSM803665 | AL832357 Homo sapi |
| 2 | 385 | 100.0 | 7928 | 6 | AX039412 | AX039412 Sequence |
| 3 | 385 | 100.0 | 20435 | 9 | HSJ2535 | AJ002635 Homo sapi |
| 4 | 377 | 97.9 | 628 | 6 | AX308286 | AX308286 Sequence |
| 5 | 274.8 | 71.4 | 4171 | 10 | BC060226 | BC060226 Mus muscu |
| 6 | 204.8 | 53.2 | 23123 | 9 | HSJ14908 | AJ314908 Homo sapi |
| 7 | 204.8 | 53.2 | 135964 | 9 | AL353593 | AL353593 Human DNA |
| 8 | 180.8 | 47.0 | 164766 | 2 | AC026657 | AC026657 Homo sapi |
| 9 | 171.4 | 44.5 | 250087 | 2 | AC096931 | AC096931 Rattus no |
| 10 | 171.4 | 44.5 | 260988 | 2 | AC099089 | AC099089 Rattus no |
| 11 | 163.4 | 42.4 | 211829 | 10 | AL645854 | AL645854 Mouse DNA |
| 12 | 93.8 | 24.4 | 174612 | 2 | AC023889 | AC023889 Homo sapi |
| 13 | 85 | 22.1 | 98613 | 2 | AC139630 | AC139630 Takifugu |
| 14 | 48 | 12.5 | 349981 | 1 | BX572602 | BX572602 Rhodopsu |
| 15 | 47.4 | 12.3 | 300425 | 1 | AP005022 | AP005022 Streptomy |
| 16 | 43.2 | 11.2 | 296500 | 1 | SC0939128 | AL939128 Streptomy |
| 17 | 42.6 | 11.1 | 4180 | 9 | HUMB94 | M92357 Homo sapien |
| 18 | 42.2 | 11.0 | 1553 | 8 | TAB276509 | AJ276509 Tricium |
| 19 | 42.2 | 11.0 | 191996 | 9 | AC092275 | AC092275 Homo sapi |
| 20 | 41.4 | 10.8 | 1803 | 6 | BD276572 | BD276572 EXTRACELL |
| 21 | 41.4 | 10.8 | 1803 | 6 | AX048206 | AX048206 Sequence |
| 22 | 41 | 10.6 | 169824 | 2 | AP001854 | AP001854 Homo sapi |
| 23 | 40.6 | 10.5 | 188462 | 9 | AL592301 | AL592301 Human DNA |
| 24 | 40.4 | 10.5 | 1078 | 1 | PATRP1G | X51868 Pseudomonas |
| 25 | 40.4 | 10.5 | 19372 | 2 | AE004443 | AE004443 Pseudomon |
| 26 | 40.2 | 10.4 | 145577 | 2 | AC145762 | AC145762 Sus scrof |
| 27 | 40.2 | 10.4 | 194809 | 4 | AC091506 | AC091506 Sus scrof |
| 28 | 40.2 | 10.4 | 292100 | 1 | SC0939121 | AL939121 Streptomy |
| 29 | 40.2 | 10.4 | 348525 | 1 | BX640428 | BX640428 Bordetell |
| 30 | 40.2 | 10.4 | 348642 | 1 | BX640446 | BX640446 Bordetell |
| 31 | 40.2 | 10.4 | 348997 | 1 | BX640427 | BX640427 Bordetell |
| 32 | 40 | 10.4 | 299800 | 1 | AP005040 | AP005040 Streptomy |
| 33 | 39.8 | 10.3 | 125020 | 9 | AF429315 | AF429315 Homo sapi |
| 34 | 39.6 | 10.3 | 9199 | 1 | AY256972 | AY256972 Pseudomon |
| 35 | 39.6 | 10.3 | 151159 | 2 | AC145010 | AC145010 Sus scrof |
| 36 | 39.6 | 10.3 | 170416 | 2 | AC145011 | AC145011 Sus scrof |
| 37 | 39.4 | 10.2 | 1341 | 9 | BC006223 | BC006223 Homo sapi |
| 38 | 39.4 | 10.2 | 2040 | 6 | AX747647 | AX747647 Sequence |
| 39 | 39.4 | 10.2 | 2040 | 9 | AK092842 | AK092842 Homo sapi |
| 40 | 39.4 | 10.2 | 123580 | 1 | AF263912 | AF263912 Streptomy |
| 41 | 39.4 | 10.2 | 125401 | 6 | AX211739 | AX211739 Sequence |
| 42 | 39.4 | 10.2 | 182662 | 9 | AC023830 | AC023830 Homo sapi |
| 43 | 39.4 | 10.2 | 207961 | 2 | AC013282 | AC013282 Homo sapi |
| 44 | 39.4 | 10.2 | 222643 | 2 | AC146802 | AC146802 Silurana |
| 45 | 38.8 | 10.1 | 3001 | 5 | LFL308113 | AJ308113 Lampetra |

ALIGNMENTS

RESULT 1
LOCUS HSM803665
DEFINITION Homo sapiens mRNA; cDNA DKFP451F056 (from clone DKFP451F056).
ACCESSION AL832357
VERSION AL832357.1 GI:21732919
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4901)
AUTHORS Koehrer,K., Beyer,A., Wewes,H.W., Weill,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Wiemann,S.
TITLE Direct Submission

JOURNAL Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@kfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the
Heinrich-Heine-University, Dueseldorf/Germany) within the cDNA
sequencing consortium of the German Genome Project. This clone
(DKFZ451F056) is available at the RZPD in Berlin. Please contact
the RZPD; Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clones@rzd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.

FEATURES
Location/Qualifiers
1..4901
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp451F056"
/db_xref="taxon:9606"
/clone="DKFZp451F056"
/tissue types="human skeletal muscle"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
polyA_signal 4828..4833
polyA_site 4847

ORIGIN
Query Match 100.0%; Score 385; DB 9; Length 4901;
Best Local Similarity 100.0%; Pred. No. 2e-71;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACCTGCGCCCTAGGGGCTGA 60
Db 1281 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACCTGCGCCCTAGGGGCTGA 1340
Qy 61 GCAGGATGCCATCAGCTGCGGAGGAGCCAGTGTGAGAGTCTTGATGCACGCCACCC 120
Db 1341 GCAGGATGCCATCAGCTGCGGAGGAGCCAGTGTGAGAGTCTTGATGCACGCCACCC 1400
Qy 121 ACTCGCTGGCTGTCGCCGCCAAGCCCAAGTCCAGCCCTCAGCGGAGGCTGGGT 180
Db 1401 ACTCGCTGGCTGTCGCCGCCAAGCCCAAGTCCAGCCCTCAGCGGAGGCTGGGT 1460
Qy 181 GTCCACGAGCTTACCTGACAGAGGAGCTCAAGCTGTCAAGTCCAGCCCTCAGCGGAGGCTGGGT 240
Db 1461 GTCCACGAGCTTACCTGACAGAGGAGCTCAAGCTGTCAAGTCCAGCCCTCAGCGGAGGCTGGGT 1520
Qy 241 CCCTGAGTTCCTGGGAGGCTGTGTCGAAGACGAATCAAGGCAAGGCTGAGCTCTGT 300
Db 1521 CCCTGAGTTCCTGGGAGGCTGTGTCGAAGACGAATCAAGGCAAGGCTGAGCTCTGT 1580
Qy 301 GATCCAGAGCTGTGAGTTCGAGCAGGCTTCGAGGAGGCTTCGAGGAGGCTGAGTTCCTGAGAG 360
Db 1581 GATCCAGAGCTGTGAGTTCGAGCAGGCTTCGAGGAGGCTTCGAGGAGGCTGAGTTCCTGAGAG 1640
Qy 361 CCACACCTGAGCAGCCTGGAGGCG 385
Db 1641 CCACACCTGAGCAGCCTGGAGGCG 1665

RESULT 2
AX039412
LOCUS AX039412
DEFINITION Sequence 5 from Patent WO0063381.
ACCESSION AX039412
VERSION AX039412.1 GI:11229480
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Zeng, W., Stanton, L. and Kong, H.

TITLE Mammalian protein with putative function in signal transduction
JOURNAL Patent: WO 0063381-A 5 26-Oct-2000;
SCIOS INC. (US)
FEATURES
Location/Qualifiers
1..7928
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
60..7850
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC16626.1"
/db_xref="GI:11229481"
/translat:cn="MLERFPTPKVKKGSSITFSVKVGPVPTVHMLREEAERGVLMI
GPDTPGYTVASSAOHSLVLLDVGROHGYTCTAGNAAGALCSALHVSGLPKVEE
OEKVEALISTFLOGTTCAISAOGLTASADLGOKKEPLAALKAALHSLAEVGT
EEFLQKLTSTENWISAKITQAKLOVPGSDSDSKTSPASPHGSRSPSSIOBSS
ESDGDARGEIDFIVVTADYLPGLASQDAITREGQIVSVLDAAPLRLWLYTKPK
SSPQRCQWSPAYDLRLKLSPEWGAEEAEFFGEVSEDEYKARUSSVIOELLSAQ
AFVSEQLQSHHLQHLRCPHVIAGQKAVFRNRDIGHFHSFVLOELQCDTD
DDVAMCFIKQAAFEQVLEFLVGRVQAEVSVTAIOEFYKAEALLAGADPSQPP
PPOHYLEQVORVORVQALLKELIRNARNONCALLQOAVAVVSAALPOAENKLHV
SLMENYPTLEALGEP:ROGHFTVMEGACGARMKGNHSHVFLFRNHLVCKPRDS
RTDVSVFRNMYKLSIDANDQVEDGRAFVWQEREDSVRYLLOAATLTKSSW
KEICGOORLALPVRPFEEELADCTAELGTVKLACRVGTGTPRPIVSWYKDGKAV
QVDPHHLIEDPDGSCALILDSLTPVDSDGQIMCFAAAGNCSTLGLKILVQVPPRVN
KVRASPFVEGEDAQTCTICAPYQIRWYKDGALLTGNKFTLSEGRGLLVIR
AASDELGLYCELVNRLGASAAELRIOSPMLOAQOECHREQLVAAVEDTLEAD
QEVTSVLRLKPKAPGPGTDLTPGQPCPRGAPALQETGSPQVTVTSEAPVPRV
POILLHGSPQEPPEAIARAQSWTVPIRMEGAWEAGCTGELLWDVSHVVRVLTQTY
TYQALDHTHTRPSNMQVITIEDVCAQTGTGTAQFEALLEGDPQSVTWKQDSVLVSTR
LSQOQEGTYSVLVRHVASKDAGYVTCIAQTGQVLCCKAELLVLDGNEPDSKOSH
RKJHSFYVEKEIGRFGVFKVQKHKKILCAAFIPRSTRRAQATREKIDILAA
LSHPLVTGLDQFETRLILILESSSELRLYRKGVVTEAVKYVIOQFVEGLH
YCOMPVSEADSIWAMGVISLSTCSSPFAGSDRATLNLVLEGRVSWSPMAHLSE
DAKDFIKATLQAPARPSAOCILSHWPKLMPABEAHFINTKOLKFLRLARWORS
LMSYKSLVWRSIPELLRPPDSPSLGVABHLCBDTGGSSSSSSSSSNEIAPARAKS
LPPSPVGHPLHPRGFRFSASLPEEAESERSTEAPAPAPGAPGAPAAAGCVP
HSVTRSLFYQAGSEPHGALAPGRHPPRRHLLKGGYIAGALGRLEPFLMEHRLV
EEEAERQATLLAKSPFETALPASPGLHAPGHSHSLEHDSPTSPSPSSACGEE
QRLSPASGGAPIDMGHPQSGKLPSTGHPGTAPQERFSPSPMGQAPFCHQKQ
SAQEGISLPKPKPGKSPGSAOASSSOVSLLRVGSSOVGTGTPGLDAEGMTQAE
GPGDISLPKPKPGKSPGSAOASSSOVSLLRVGSSOVGTGTPGLDAEGMTQAE
DLSTSTLQRPQATWPKFSLGGRGYAGVAGYGTFAFGDAGMLGOGPKWART
WAVSQSEEEQEAASQSEQEAASFLQVSAFPVPEVGRAPTRSPSPETPK
DIGQSVLVQIRLSDGDAADTISLISEVDPAFLNLSLDYDKILPFEFMIKPKPK
SAQEPSPMAEEELAEFPFTWPMGELGHAGLEITEESDVLALAAVGRKKK
WSPSRSLFFPGHPLPDLPAELGRVRYKSHSRILKGRPELEKEGPPRKP
GLAGFLSLGKSWDRAPTLFRLSDTGLQSVTLACQVSAQAPAAQKXGDALE
SSSVLISATLKNFOLLTIIVVAEDLGVTCVSNALGTVTTVTGLVRKABRSPPC
PDIGVANDGVLYWKPVSYPVTYVQCSLEGGSWTTLASDIFDCQYTLKLSRGG
SPRHLVILELCSGPELLCLAEASYSSEVKDYLWQMLSATQYLHNLHLDLRS
ENMIITEYNLLKVDNLGNLSQELVKDFKDYLETWAPLELQGGVAPQTDIWA
IGVTAFMLGAEYVPSSEGARDLQRLKGLRSLRCYAGLSGVAFLRSTQIAQW
GRPCASSCLQCPWLITESPACSRPAPVFTFTARLVRVFRNREKRLLYKRNLAQVR

ORIGIN
Query Match 100.0%; Score 385; DB 6; Length 7928;
Best Local Similarity 100.0%; Pred. No. 1.9e-71;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACCTGCGCCCTAGGGGCTGA 60
Db 734 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACCTGCGCCCTAGGGGCTGA 793
Qy 61 GCAGGATGCCATCAGCTGCGGAGGAGCCAGTGTGAGAGTCTTGATGCACGCCACCC 120
Db 794 GCAGGATGCCATCAGCTGCGGAGGAGCCAGTGTGAGAGTCTTGATGCACGCCACCC 853

TRRRSADSCQSSTTTELSTQTQVNFVGTGTTLHICPDRDGDAAQP"

| | | | |
|--|---|-----------|---|
| exon | 15192..15303 /gene="OBSCN" /number=85 | exon | 15192..15303 /gene="OBSCN" /number=85 |
| intron | 15304..18393 /gene="OBSCN" /number=85 | intron | 15304..18393 /gene="OBSCN" /number=85 |
| exon | 18394..18436 /gene="OBSCN" /number=86 | exon | 18394..18436 /gene="OBSCN" /number=86 |
| intron | 18437..21834 /gene="OBSCN" /number=86 | intron | 18437..21834 /gene="OBSCN" /number=86 |
| exon | 21835..23123 /gene="OBSCN" /number=87 | exon | 21835..23123 /gene="OBSCN" /number=87 |
| ORIGIN | | | |
| Query Match 53.2%; Score 204.8; DB 9; Length 23123; | | | |
| Best Local Similarity 99.0%; Pred. No. 1.8e-33; | | | |
| Matches 206; Conservative 0; Mismatches 2; Indels 0; Gaps 0; | | | |
| QY | 9 AGATCTTTGACATCTACGTGCTACCGCTGACTACCTGCTCCCTAGGCGCTGACGAGT 68 | QY | 9 AGATCTTTGACATCTACGTGCTACCGCTGACTACCTGCTCCCTAGGCGCTGACGAGT 68 |
| Db | 210 AGATCTTTGACATCTACGTGCTACCGCTGACTACCTGCTCCCTAGGCGCTGACGAGT 269 | Db | 210 AGATCTTTGACATCTACGTGCTACCGCTGACTACCTGCTCCCTAGGCGCTGACGAGT 269 |
| QY | 69 CCATCAGCTGCGGGAAGCCAGTATGTGGAGTCTCTGGATGCGAGCCACCCACTGGCT 128 | QY | 69 CCATCAGCTGCGGGAAGCCAGTATGTGGAGTCTCTGGATGCGAGCCACCCACTGGCT 128 |
| Db | 270 CCATCAGCTGCGGGAAGCCAGTATGTGGAGTCTCTGGATGCGAGCCACCCACTGGCT 329 | Db | 270 CCATCAGCTGCGGGAAGCCAGTATGTGGAGTCTCTGGATGCGAGCCACCCACTGGCT 329 |
| QY | 129 GCGTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGGCGGCTGGGTGTACCCAG 188 | QY | 129 GCGTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGGCGGCTGGGTGTACCCAG 188 |
| Db | 330 GCGTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGGCGGCTGGGTGTACCCAG 389 | Db | 330 GCGTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGGCGGCTGGGTGTACCCAG 389 |
| QY | 189 CCTACCTGGACAGGAGGCTCAAGCTGTC 216 | QY | 189 CCTACCTGGACAGGAGGCTCAAGCTGTC 216 |
| Db | 390 CCTACCTGGACAGGAGGCTCAAGCTGTC 417 | Db | 390 CCTACCTGGACAGGAGGCTCAAGCTGTC 417 |
| RESULT 7 | | | |
| AL353593 | Human DNA sequence from clone RP5-1139B12 on chromosome 1q42.1-43, complete sequence. | AL353593 | Human DNA sequence from clone RP5-1139B12 on chromosome 1q42.1-43, complete sequence. |
| ACCESSION | AL353593 | ACCESSION | AL353593 |
| VERSION | AL353593.33 GI:18673899 | VERSION | AL353593.33 GI:18673899 |
| KEYWORDS | HTG | KEYWORDS | HTG |
| SOURCE | Homo sapiens (human) | SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens | ORGANISM | Homo sapiens |
| REFERENCE | 1 (bases 1 to 135964) | REFERENCE | 1 (bases 1 to 135964) |
| AUTHORS | Almeida J. | AUTHORS | Almeida J. |
| TITLE | Direct Submission | TITLE | Direct Submission |
| JOURNAL | Submitted (13-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk On Feb 14, 2002 this sequence version replaced gi:17977879. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. | JOURNAL | Submitted (13-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk On Feb 14, 2002 this sequence version replaced gi:17977879. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. |
| COMMENT | This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP | COMMENT | This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP |

Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleneh, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindester, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Sma's, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Staible, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 250087)
 Worley, K.C.
 Direct Submission
 Submitted (03-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 250087)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On May 10, 2003 this sequence version replaced gi:23665228.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GFYV
 Center clone name: CH230-96E21
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 239508 bases at least Q40
 Consensus quality: 242433 bases at least Q30
 Consensus quality: 244682 bases at least Q20
 Estimated insert size: 248556; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence

* as soon as it is available and the accession number will be preserved.

1 4351: contig of 4351 bp in length
 4352 4451: gap of unknown length
 235250: contig of 230799 bp in length
 235350: gap of unknown length
 235351 236367: contig of 1017 bp in length
 236368 236457: gap of unknown length
 237481: contig of 1014 bp in length
 237581: gap of unknown length
 238973: contig of 1292 bp in length
 238974 238973: gap of unknown length
 240258: contig of 1285 bp in length
 240259 240358: gap of unknown length
 241550: contig of 1292 bp in length
 241551 243880: contig of 2130 bp in length
 243881 243980: gap of unknown length
 246745: contig of 2765 bp in length
 246746 246845: gap of unknown length
 250087: contig of 3242 bp in length.

FEATURES
 Location/Qualifiers
 1..250087
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-96E21"
 1..1462
 /note="wgs end extension
 clone_end:Sp6"
 418..1115
 /note="clone boundary
 clone_end:Sp6
 site:ECORI
 end_sequence:BH273653"
 complement(227830..228525)
 /note="clone boundary
 clone_end:T7
 site:ECORI
 end_sequence:BH273651"
 ORIGIN
 Query Match 44.5%; Score 171.4; DB 2; Length 250087;
 Best Local Similarity 89.8%; Pred. No. 1.3e-26;
 Matches 184; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
 QY 9 AGATCTTTGACATCTACGTGGTCAACGCTGATCTGAGGTCAGCCCTAGGCGGTGAGCAGGATG 68
 DB 241387 AGATCTTTGACATCTACGTGGTCAACGCTGATCTGAGGTCAGCCCTAGGCGGTGAGCAGGATG 241328
 QY 69 CCATCAGCTGCGGAGGCCAGTATGTGGAGGTCTGTGATGACAGCCACCCACTGCGCT 128
 DB 241327 CCATCAGCTGAGAGGCCAGTATGTGGAGGTCTGTGAGTCTCCCTGACCTGAGGCTG 241268
 QY 129 GCGTTCTCGCAGCCAGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTCAACCAAG 188
 DB 241267 GCGTTCTAGCAGCCAGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTCAACCTG 241208
 QY 189 CCTACCTGACAGGAGGCTCAAGCT 213
 DB 241207 CCTACCTGACAGGAGGCTCAAGCT 241183
 RESULT 10
 AC099089/c
 LOCUS
 Rattus norvegicus clone CH230-154E3, WORKING DRAFT SEQUENCE.
 AC099089
 AC099089, 5 GI:30522701
 HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 Rattus norvegicus (Norway rat)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

| SOURCE | Mus musculus (house mouse) |
|---------------------------|--|
| ORGANISM | Mus musculus |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |
| AUTHORS | Smith M. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerquest@sanger.ac.uk |
| COMMENT | On Apr 7, 2002 this sequence version replaced gi.19699553. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TRSMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-192P17 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBACe3.6. |
| FEATURES | <p>Location/Qualifiers</p> <p>1..211829</p> <p>/organism="Mus musculus"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:10090"</p> <p>/chromosome="11"</p> <p>/clone="RP23-192P17"</p> <p>/clone_lib="RPCI-23"</p> |
| ORIGIN | |
| Query Match | 42.4%; Score 163.4; DB 10; Length 211829; |
| Best Local Similarity | 87.3%; Pred.No. 6.7e-25; |
| Matches 179; Conservative | 0; Mismatches 26; Indels 0; Gaps 0; |
| Qy | 9 AGATCTTTGACATCTACGTGTCACCGCTGACTACTCTGCCCTAGCGGCTGACGAGATG 68 |
| Db | 15366 AGATCTTTGACATCTATGTGGTCACACTGACTATCTGCCACTGGGAGCTGACGAGATG 15300 |
| Qy | 69 CCATCAAGCTCGCGGAAGGCCAGTATGTGGAGGTCTCTGGATGTCAGCCACCCCACTGGCGT 128 |
| Db | 15306 CCATCAATTCGTGAGAGAGGCCAGTATGTGGAGGTCTCTGACTCTGCCATCCCTGGCT 1524 |
| Qy | 129 GGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCACGGCAGGGCTGGGTGTCCACG 188 |
| Db | 15246 GGCTTGTCCGGAACAAGCCACCAAAATCCAGTCTCTCCAGGCGGGCTGGGTGTCCCTG 1518 |
| Qy | 189 CCTACCTGGACAGGAGGCTCAAGCT 213 |
| Db | 15186 CCTACCTGGATAGAGGCTCAAGGT 15162 |
| RESULT 12 | |
| AC023889/c | |
| LOCUS | 174612 bp DNA linear HTG 07-JUL-2000 |
| DEFINITION | Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT |
| SEQUENCE | 13 unordered pieces. |
| AC023889 | |
| ACCESSION | AC023889.3 GI:8969253 |
| VERSION | HTG; HTGS PHASE1; HTGS_DRAFT. |
| KEYWORDS | Homo sapiens (human) |
| SOURCE | |

```

/misc_feature 1. 2478 /chromosome="1"
/misc_feature 2579. 6948 /clone="RP11-661B12"
/misc_feature 7049. 9373 /note="assembly_name:Contig4"
/misc_feature 9474. 13303 /note="assembly_name:Contig5"
/misc_feature 13404. 18270 /note="assembly_name:Contig6"
/misc_feature 18371. 25589 /note="assembly_name:Contig7"
/misc_feature 25690. 34815 /note="assembly_name:Contig8"
/misc_feature 34916. 45742 /note="assembly_name:Contig9"
/misc_feature 45943. 62987 /note="assembly_name:Contig10"
/misc_feature 63088. 87584 /note="assembly_name:Contig11"
/misc_feature 87685. 114765 /note="assembly_name:Contig12"
/misc_feature 114866. 142334 /note="assembly_name:Contig13"
/misc_feature 142435. 174612 /note="assembly_name:Contig14"
/misc_feature 174612 /note="assembly_name:Contig15"
/misc_feature 174612 /note="assembly_name:Contig16"

Query Match 24.4% Score 93.8; DB 2; Length 174612;
Best Local Similarity 97.9%; Pred. No. 3.5e-10;
Matches 95; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 289 GCTGAGCTCTGTGATCCAGAGCTGTGAGTTCTGAGCAGCGCTTCTGTGAGAGGTGCA 348
Db 86938 GCCAGCTCTGTGATCCAGAGCTGTGAGTTCTGAGCAGCGCTTCTGTGAGAGGTGCA 86779

Qy 349 GTTCTCAGAGCCACCTGAGCAGCCTGAGCGC 385
Db 86778 GTTCTCAGAGCCACCTGAGCAGCCTGAGCGC 86742

RESULT 13
AC139630 LOCUS 98613 bp DNA linear HTG 08-MAR-2003
DEFINITION Takifugu rubripes clone 221D8, WORKING DRAFT SEQUENCE, 6 ordered
pieces.
ACCESSION AC139630
VERSION AC139630.2 GI:28882132
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 98613)
AUTHORS Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Cariga, K., Coleman, B., Engle, J., Grant, S., Guan, X., Gupta, J.,
Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,
Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,
Marquies, E.H., Masello, C., Maskeri, B., McDowell, J.,
Paquigan, C., Pearson, R., Portnoy, M.E., Prasad, A.,
Reddix-Dugue, N., Schandler, K., Schueler, M.G., Sison, C.,
Stratipop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,
Wetherby, K.D., Wiggins, L., Young, A., and Green, E.D.
TITLE NISC Comparative Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 98613)

```

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Green, E.D.
Direct Submission
Submitted (07-FEB-2003) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 98613)
Green, E.D.
Direct Submission
Submitted (08-MAR-2003) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Mar 8, 2003 this sequence version replaced gi:28469380.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nih.gov
----- Project Information
Center project name: egw
Center clone name: 221D08

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 97841 bases at least Q40
Consensus quality: 98004 bases at least Q30
Insert size: 85000; agarose-fp
Insert size: 98113; sum-of-contigs
Quality coverage: 11.51x in Q20 bases; agarose-fp
Quality coverage: 9.97x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 28467: contig of 28467 bp in length
* 28468 28567: gap of unknown length
* 28568 31135: contig of 2568 bp in length
* 31136 31235: gap of unknown length
* 31236 41487: contig of 10252 bp in length
* 41488 41587: contig of unknown length
* 41588 74365: contig of 32778 bp in length
* 74366 74465: gap of unknown length
* 74466 95541: contig of 22076 bp in length
* 95542 96642: gap of unknown length
* 96642 96613: contig of 1972 bp in length.

Location/Qualifiers
1. 98613
/organism="Takifugu rubripes"
/mol_type="genomic DNA"
/db_xref="taxon:31033"
/clone="221D8"
/clone_lib="Incyte Genomics"
1. 28467
/note="assembly_fragment"
clone_end:T7

misc_feature

Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.
Complete genome sequence and comparative analysis of the industrial
microorganism Streptomyces avermitilis
Nat. Biotechnol. 21 (5), 526-531 (2003)
22608306
PUBMED
12692562
3 (bases 1 to 300425)
Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osuno,T.,
Kushida,N., Director-General of Biotechnology Center, Shiba,T.,
Sakaki,Y. and Hattori,M.
Direct Submission
Submitted (29-WAR-2002) Director-General of Biotechnology Center,
National Institute of Technology and Evaluation, Biotechnology
Center; 2-49-10 Nishinara, Shiba-ku, Tokyo 151-0066, Japan
(E-mail:bionite.go.jp. URL:http://www.bio.nite.go.jp/,
Tel:81-3-3481-1933, Fax:81-3-3481-8424)
This work was done in collaboration with Haruo Ikeda(*1), Jun
Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi
Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi
Osuno(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi
Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7)
and Satoshi Omura(*1,*3).
Final finishing process and all annotation were done by H. Ikeda
and J. Ishikawa.
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
http://avermitilis.ls.kitasato-u.ac.jp.
Location/Qualifiers
1. 300425
/organism="Streptomyces avermitilis MA-4680"
/mol_type="genomic DNA"
/strain="MA-4680"
/db_xref="taxon:227882"
/note="This strain is also named as strain: ATCC 31267,
NCIMB 12904 of NRRL 8165.-synonym: Streptomyces
avermectinus"
complement(382. .645)
/note="SAV258"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAC67967.1"
/db_xref="GI:29603895"
/translation="WMNFALAPEVSVTDTTHGVLLDRTGRYLLNGTALRALRL
DGTSPQGVASLCRFPDAGTAEQVDNLHAIRTAQVAAS"
complement(642. .2552)
/note="SAV259"
/codon_start=1
/transl_table=1
/product="putative membrane-associated Zn-dependent
protease"
/protein_id="BAC67968.1"
/db_xref="GI:29603896"
/translation="MPAVRPLHLLTLVEGTGMEFLVLPDPAGAAALALMPTAGPPPP
THVIEHSGRPWVGNCNEEDVMYTPRPVALLGRTADADVLTERLDVRSVRDL
DGLAATAGSFLVSVNGEVRVQGLTSACQIYATVAGTVLAADSPQALADRIAGD
IDEDQALQALAEVATCTAGQGVVSADSLGSDTSLCFLADRDHAHLVTVHYH
RPLNGMDIRQWAPQAALPAGRHLLVPGQAPCPYAPLAHAVHDREGPLTFVRSRA
QALRGSDDRWARQAALPAGRHLLVPGQAPCPYAPLAHAVHDREGPLTFVRSRA
TLEHLAALAAAGRRRLQIGCGDELPHSPMALHALVRHRLPAAALPHLRALSRMVR
PLTATTLAASPYTQWLACSNFPADRDHWGADWEPLKLPNATPHAAATVRVR
QLLAALATPGLATLPVQHEMIRLTIRINGNATRNNSHIGHTGVTHAPTTDDRVLE
AILAVRDMRVALGQFPMQGTALRSIVPAGLLRTTKGDASBLVAGLRRRRRELLK
LMDDSHLVRFMGVDVADLVIRGLHADTRTLMPELDDSTLACELMLRTLPADTARA
VPFSPRTVKEI"
3174. .3572

/note="SAV260"
 /codon_start=1
 /transl_table=11
 /product="putative transposase"
 /protein_id="BAC67969.1"
 /db_xref="GI:29603897"
 /translation="MGRAGPVAFLTDEQRNYCTENOVDDGQLAGVFLDDRARRA
 MTKGRASQUGYAVQUGTVFLGTFLDNFEDAPAEVAVVAEQLGHPGVSVLACYGNER
 TRWDHQIAIKDAYGYRDLKGDAWKKLRDGD"
 3604..3834
 /note="SAV261"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC67970.1"
 /db_xref="GI:29603898"
 /translation="MLGQGYVPLDLMIGESQMSRSDALRGLRALMMVMVIAIEVPV
 YLWMLVEVSYLTGQLPRGTTEBDGKPRADTD"
 4027..4500
 /note="SAV262"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC67971.1"
 /db_xref="GI:29603899"
 /translation="MQADAMAVAMELOQAAANRLIWEGPAERGETFLTLIAFAFG
 AAPARIAGGTDVDSGLVGFQGAELLISRRVASKQTAAVREPLAAQATAAAALPMRCS
 APAVAAETEQLLNALADVENTARLEAALEAFGRVAVATPPRLSWARREAVSSG"
 complement(4811..6445)
 /note="IS6 family; completely same as SAV7545 and SAV75480
 SAV263"
 /codon_start=1
 /transl_table=11
 /product="putative IS6 family transposase"
 /protein_id="BAC67972.1"
 /db_xref="GI:29603900"
 /translation="MILAECGELLKTMFPLEGVLEEVSPESGVLIHVARTVESVPLP
 PCPCATPSVRHSGYORLADGAVGRQVSIELTVRRLCDQRCARVTFABQDGL
 TVRYGRTPOLRCLLSAIVALAGRERLAARLPVSRITLLIALVMALPDPYATP
 RVLGDFEATRKHGKGYTVLDVCETHAPIDLLPDSRATFAAMLDHPGVETICRDR
 CAFADGATGPADAVQADLWHLWHNLATVKSIVKSHSCLREPALASPPVELVHP
 PLSHAGRLAERAPHAADVLDLGGGLTIRAVARRLELSNTVRYVARAVTWEELATG
 RWQLPSTLDPKYSLHQRWHEGHTSAKLHAELRERGFTGSYVVYRDLVRRFRRTSG
 DRPPFAPPGVKVATGMITNPRMNDQDOOKAILARCPELEAATGCHVRSFAPMA
 IRSARLPFETARADEHNLGDFADGLDLADLILGLUSTEWSSGCVREGRVTDIKL
 LKRWAGRAGLPLLRKVLVAADRRQHRVTNOTAH"
 complement(6592..8902)
 /note="SAV264"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC67973.1"
 /db_xref="GI:29603901"
 /translation="WTPTPTPSTSPSPSQCGHGTTPEDPAGRCGIRVSGHTACIAL
 HLSADRDVAYLHSPGNVNDHRTGSFQDLDDALLDLPDTPDRLGVAPRSAT
 FQDVRFSATFQKQADPGSATFOGNAGFLATPEGPAGFSARFEGDAGFGSARFG
 DASVFSAITFEGDASFVATFEGDASFVSAITFQSAASLGWAAFGQDVKEESATQGDVA
 FGGADTFDGAFAFLAERNARFGARFGISFLWTFHGDADFSRAIFRGAOSFV
 ATFQSDTLEAATSDHTRFATFQCDGFVEATFEGDTSFVSATLQSNALFGLVTF
 ERDAKFSSIFQGAATFVSATFECDAFEQVGFERSASGLGLVCAGRVKLSGAVFGCP
 LTFAPATLCECRTRVSTAEELRLVATVDFAHVEYPLTTAAADPVEIPLHNGRPV
 AEQSLIADPDEVIRASLRGVDAHLVSDVLSGCLTGTVHLDQIRLGGTCSFASV
 PRKHWRNRRLRTRQRRALAEQFWPRAVDTPVPGWNAVAGAVHAGPAQLAPVYRA
 LRKPFDRSRNFGAADTYIGEMENRHRDRTGSTRERGLLHGHWLLSGYGLRASRALG
 WLAAMALTVLLMGFGPKDPSKQVATGVPDGGKVLVLEIDKGLDQNPQNTDRGFISE
 RFKALCTILNSVVRSSFGDPLATTAGTYIEMASRVTFPVLGLLGLAVLRNRVKR"
 9294..9707
 /note="SAV265"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC67974.1"
 /db_xref="GI:29603902"

/translation="MGTRLRRADFSDRCERKLTOKVPVSRPQGVQGLTNTYLSP
 TEYDLLASLPAAVLKTRFSVPPLGVDFVDFQPLQGLVLAEEFTTDEETQSFIPPSEC
 VAETDDARFTGGRLVETSRQELLGLAEYGLRPQ"
 complement(9955..10614)
 /note="SAV266"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC67975.1"
 /db_xref="GI:29603903"
 /translation="MPHVLRKRESSCYLNRLADANRLPVQFRPILGHNHRRD
 DNRGRWPTTCLRTAEALSHAI PALAVIGEPAAENLQSLSGQAVEVLRFP
 ACRCMARFGHVLVRTIPPEAVCLRHQRWLLGDEQYRLHALPELQANRRHCLV
 RRTASILAREYAVARDCLRDWFQDGKPSALRQWNKRIQLGEDPVGDETPVAPSP
 G"
 10830..11180
 /note="SAV267"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC67976.1"
 /db_xref="GI:29603904"
 /translation="MLEDDPAHGVLDCTNVLHETGDEKPAEHAEEFVLPHLEH
 RVLVQVARASRSLEIVIDDSPQRAIVERGPWALWDEYESGTVPPQGGIRLCLSLH
 AKGNRWPLSFTTC"
 complement(11530..11979)
 /note="SAV268"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC67977.1"
 /db_xref="GI:29603905"

CDS

Query Match 12.3%; Score 47.4; DB 1; Length 300425;
 Best Local Similarity 49.0%; Pred. No. 2;
 Matches 126; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
 QY 18 ACATCTACGTGGTCACCGCTGACTACCTGCCCTAGGGGTGAGCAGGATGCCATCAGC 77
 Db 171375 ACATCGACCTCGAGACCGCGACCGACTGCCGTGCGGCAGTTACGTGCTTACCTCAGG 171434
 QY 78 TCGGGAAGGCCAGTATGTGGAGTCTCTGGATGCAGCCACCCACTGCCGTGGCTTGTC 137
 Db 171435 TCTGGGAACGCCAGTCCACCGAGTCCGAGACCCCTCGCTCCACGAAACCGCGCTCGGC 171494
 QY 138 GCACCAAGCCACCAAGTCCAGCCCTCAGCGAGGGCTGGGTGTCCAGAGCTACCTGG 197
 Db 171495 TGCACCAAGCCGACTCGCGAGCGCGCCAGGTGTTGTGGCAGGTCCGCGCACTGCCGC 171554
 QY 198 ACAGGAGGCTCAAGCTGTACCTGAGTGGGGGGCGCTGAGGCCCTGAGTTCCCTGGGG 257
 Db 171555 TCAGAGCGCGGACAGCCCGTGCACCGGATCGCCCGGAGCAGCACCGGACCGGGGCA 171614
 QY 258 AGCTGTGTCTGAAGAC 274
 Db 171615 CGGCGTCTCTCGACGTC 171631

Search completed: September 19, 2004, 19:27:13
 Job time : 1576.62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 14:16:08 ; Search time 1570.02 Seconds
(without alignments)
10628.572 Million cell updates/sec

Title: US-10-077-130-4 COPY 10286 10670

Perfect score:

Sequence: 1 acatagcctgaggcaggacg.....cctcagccagcgtcaccatc 385

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : GenEmbl:*

1: gb_ba: *
2: gb_hug: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|-------|--------|----|-----------|---------------------|
| | | | | | | | |
| C | 1 | 385 | 100.0 | 4789 | 9 | AB046776 | AB046776 Homo sapi |
| | 2 | 385 | 100.0 | 20435 | 9 | HGAJ7535 | AJ002535 Homo sapi |
| | 3 | 261 | 67.8 | 34667 | 9 | HGA314905 | AJ1314905 Homo sapi |
| | 4 | 281 | 67.8 | 135964 | 9 | AL353593 | AL353593 Human DNA |
| | 5 | 281 | 67.8 | 164766 | 2 | AC026657 | AC026657 Homo sapi |
| | 6 | 261 | 67.8 | 174612 | 2 | AC023889 | AC023889 Homo sapi |
| | 7 | 253.6 | 65.9 | 352 | 6 | AX069754 | AX069754 Sequence |
| | 8 | 196 | 50.9 | 211829 | 10 | AL645854 | AL645854 Mouse DNA |
| | 9 | 192.2 | 49.9 | 164766 | 2 | AC026657 | AC026657 Homo sapi |
| | 10 | 180.6 | 49.5 | 2155 | 6 | AX463538 | AX463538 Sequence |
| C | 11 | 184.2 | 47.8 | 26098 | 2 | AC099089 | AC099089 Rattus no |
| | 12 | 174.6 | 45.4 | 2295 | 9 | HSM800639 | AL080126 Homo sapi |
| | 13 | 165 | 42.9 | 621 | 11 | BV014682 | BV014682 S212P602 |
| | 14 | 71.4 | 18.5 | 98613 | 2 | AC139630 | AC139630 Takifugu |
| C | 15 | 70 | 18.2 | 2534 | 6 | AX834822 | AX834822 Sequence |
| | 16 | 70 | 18.2 | 2534 | 6 | AX097489 | AX097489 Homo sapi |
| | 17 | 70 | 18.2 | 2761 | 9 | BC043916 | BC043916 Homo sapi |
| | 18 | 55 | 14.3 | 180872 | 5 | AL844185 | AL844185 Zebrafish |
| C | 19 | 54.2 | 14.1 | 7218 | 6 | I66494 | I66494 Sequence 14 |
| | 20 | 50.4 | 13.1 | 6400 | 9 | HSA314903 | AJ1314903 Homo sapi |
| | 21 | 50 | 13.0 | 6602 | 4 | AX136512 | AX136512 Canis fam |
| | 22 | 49.4 | 12.8 | 252070 | 2 | AC098426 | AC098426 Rattus no |
| | 23 | 47.6 | 12.4 | 5382 | 9 | AB014557 | AB014557 Homo sapi |
| | 24 | 46.8 | 12.2 | 31595 | 6 | AX780060 | AX780060 Sequence |
| | 25 | 46.8 | 12.2 | 81940 | 9 | HST1IN2B | X90568 H.sapiens m |
| | 26 | 46.8 | 12.2 | 93801 | 6 | BD186121 | BD186121 Preventio |
| | 27 | 46.2 | 12.0 | 174612 | 2 | AC023889 | AC023889 Homo sapi |
| | 28 | 44.8 | 11.6 | 289 | 6 | AR162089 | AR162089 Sequence |
| C | 29 | 44.8 | 11.6 | 289 | 6 | AR166614 | AR166614 Sequence |
| | 30 | 44.6 | 11.6 | 53424 | 3 | AB055861 | AB055861 Procamb |
| | 31 | 43.6 | 11.3 | 356 | 5 | OMF517947 | AUS17947 Oncohync |
| | 32 | 43 | 11.2 | 2545 | 9 | HSA314898 | AJ1314898 Homo sapi |
| | 33 | 42.8 | 11.1 | 49721 | 2 | AC104991 | AC104991 Homo sapi |
| | 34 | 42.4 | 11.0 | 19475 | 9 | HSA314900 | AJ1314900 Homo sapi |
| | 35 | 42.4 | 11.0 | 258174 | 2 | AC079429 | AC079429 Mus muscu |
| | 36 | 42.2 | 11.0 | 371 | 5 | OMF517957 | AJ517957 Oncorync |
| | 37 | 42.2 | 11.0 | 300817 | 1 | AB016943 | AB016943 Bacteroid |
| | 38 | 42 | 10.9 | 1135 | 9 | BC026297 | BC026297 Homo sapi |
| C | 39 | 42 | 10.9 | 1841 | 9 | BC030823 | BC030823 Homo sapi |
| | 40 | 42 | 10.9 | 4833 | 9 | AKI25056 | AKI25056 Homo sapi |
| | 41 | 42 | 10.9 | 23887 | 9 | HST1IN2 | X90569 H.sapiens m |
| | 42 | 42 | 10.9 | 207377 | 2 | AC118235 | AC118235 Mus muscu |
| | 43 | 42 | 10.9 | 222707 | 2 | AC079425 | AC079425 Mus muscu |
| | 44 | 41.8 | 10.9 | 933 | 9 | BC061909 | BC061909 Homo sapi |
| | 45 | 41.8 | 10.9 | 1704 | 9 | AF035292 | AF035292 Homo sapi |
| | 46 | 41.8 | 10.9 | 1704 | 9 | AF035292 | AF035292 Homo sapi |

ATTACHMENTS

[illegible]

which code for large proteins in vitro
 DNA Res. 7 (4), 273-281 (2000)
 20450683
 10997877
 2 (bases 1 to 4789)
 Ohara, O., Nagase, T. and Kikuno, R.
 Direct Submission
 Submitted (03-AUG-2000) Osamu Ohara, Kazusa DNA Research Institute,
 Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
 292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp,
 URL: http://www.kazusa.or.jp/huge, Tel: 81-438-52-3913,
 Fax: 81-438-52-3914)

FEATURES

source

1. 4789
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="fhl18619"
 /note="vector:pBluescriptII SK plus"
 1. 4789
 /gene="KIAA1556"
 <i>1..>4789
 /gene="KIAA1556"
 /note="Start codon is not identified."
 /codon_start=3
 /product="KIAA1556 protein"
 /protein_id="BAB13382.1"
 /db_xref="GI:10047177"
 /translation="RARRVHDJLVGKTKVLEVLGEGSCFEVLSHESASDPAM
 WVGKTVSSRSQFARQGRKYLIVVREAPDAGEWFSVRLTSKASLIVRERPA
 AIIKPLDQWVAPEDVYELRCELRSAGTQPVHLKDRKAIKSKQYDVCEGTWMLVPI
 RASLKDAGEVTEVEASKSTASHVEEKANCFTELTNLQVEKGTAVFTCKTEHPA
 ATVTWRGLLELRAGKHQPSQEGTLRLTISALEKADSDTYTCDIGQASQALLVQ
 GRRVHIITLEDVDVQSGSATPCRISPNYVPVHFKLDKTPHLANELNEIDAQGG
 YHVLTLQALQSGTIFPAGQORASALURVTEKFSVRSRELTADITTEGEDTLVLC
 ETSCDIPVCKDGTIRGSAQCSHESGHAQLLITGATLQDSKCEYKCEAGACSS
 SIVVHARVPFQALXDLVELEGATLFCVLSVAAPVKYGNVLRPGDKSLR
 QEGAMLELVNRLPQDSGRYSQFGDTTATLTALPAQFIKLRKEATEGATA
 TRLCELSKAPVWRKGSSETLRDQRCYLRQDQMCCELQIRGLAWDAEYSCVGEH
 RTSASLTIRMPAHFIRLROESIGATATLRCESLKAAPVWRKGRSLSDGRHS
 LRQDGAIVCELIQGLAVADAGEYSCVGEERTSATLTVALPAKFTGEGIRNEAVEGA
 TMLWCESKVAPEWRKGPENLRDGRYILRQETRCBELQICLAWADAGEYLCVCG
 QERTSATLTVALPARFTQDLTKKASRGATATLQCELSKAPVWRKGSSETLRDGR
 YSLRQDQTCLEIQLGLAMADTGEYSCVQERTSATLTVALPIKFTGEGIRNEATE
 GATVALRCELKMAPEVWRKGHETLRDGRSLRQDQRCARCELIQIRGLVAEDAGEYLCM
 CGKERTSATLTVALPARFTGEGIRNEATEGATATLWCESLKAAPVWRKGHETLRDGR
 DRHSRQDGRSCELQIRGLAVADAGEYSCVQERTSATLTVALPARFTGEGIRNEATE
 REGATAVLOCELSKAPVWRKGSSETLRDGRYSLRQDQRCLEIQLHGLSVADTGEYS
 CVQERTSATLTVALPARFTQDLTKKASRGATATLQCELSKAPVWRKGSSETLRDGR
 DGYSILKQDGTCELIQHLDSVADAGEYSCVQERTSATLTVALPARFTGEGIRNEATE
 EMEGATATLQCELSKAPVWRKGHETLRDGRSLRQDQRCARCELIQIRGLAMADNGY
 YSCVQERTSATLTVALPARFTGEGIRNEATEGATATLQCELSKAPVWRKGPNT
 LKQDGRYSRQDQTCLEIQLGLAVADAGEYSCVQERTSATLTVALPARFTGEGIRNEATE
 NHEATEGATATLQCELSKAPVWRKGSSETLRDGRYSLRQDQRCLEIQLHGLSVADT
 GEYLCVQERTSATLTVALPARFTIDNWTNQEAREGATATLHCELSKAPVWRKGP
 ETLRQDGRHSRQENRLNFGGGCSLGSCHCSPAWTAKLDFIS"

gene

CDS

RESULT 2
 HSAJ2535
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

HSAJ2535
 Homo sapiens mRNA for obscurin (OBSCN gene).
 AJ002535
 AJ002535.1 GI:15026973
 OBSCN gene; obscurin.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

FEATURES

source

gene

CDS

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

0; Mismatches

0; Indels

0; Gaps

0; Length

4789;

Pred. No. 4.3e-78;

Score 385; DB 9;

100.0%;

ACATAGCTGAGCAGACGGGCTGTGCGAGCTCGAGATCTGTGGCTCGCTGTGGC 60

2042

ACATAGCTGAGCAGACGGGCTGTGCGAGCTGTGCGAGCTGTGCGCTGTGGC 2101

61

AGATGCTGGGAGTACTCTCTGTGTGTGTGGGAGGAGAGGACCTCTGCCACTCTCACCGT 120

2102

AGATGCTGGGAGTACTCTCTGTGTGTGTGGGAGGAGAGGACCTCTGCCACTCTCACCGT 2161

121

GAAGGCCCTGCCAGCAAGTTCACAGAGGCTGTGAGGATGAAGAGCGCGTGGAGGGGC 180

2162

GAAGGCCCTGCCAGCAAGTTCACAGAGGCTGTGAGGATGAAGAGCGCGTGGAGGGGC 2221

QY 181 CACAGCCATGTTGTGTGTGAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGGCC 240
 Db 2222 CACAGCCATGTTGTGTGTGAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGGCC 2281
 QY 241 CGAAGACCTCAGAGATGGGACAGATACATCTCAGGAGGAGGAGGACAGAGTGTGAGCT 300
 Db 2282 CGAAGACCTCAGAGATGGGACAGATACATCTCAGGAGGAGGAGGACAGAGTGTGAGCT 2341
 QY 301 GCAGATCTGTGGCCATGGCCATGGCGAGCGCCGGGAGTACTTGTGTGTGGGCGCAGA 360
 Db 2342 GCAGATCTGTGGCCATGGCCATGGCGAGCGCCGGGAGTACTTGTGTGTGGGCGCAGA 2401
 QY 361 GAGGACCTCAGCCACGCTCACCATC 385
 Db 2402 GAGGACCTCAGCCACGCTCACCATC 2426

HSAJ2535
 Homo sapiens mRNA for obscurin (OBSCN gene).
 AJ002535
 AJ002535.1 GI:15026973
 OBSCN gene; obscurin.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Young, P., Ehler, B. and Gautel, M.
 Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor
 protein involved in sarcomere assembly
 J. Cell Biol. 154 (1), 123-136 (2001)

2 (bases 1 to 20435)
 PubMed 11448995

Gautel, M.S.
 Direct Submission

Submitted (31-OCT-1997) Gautel M.S., Structural Biology Division,
 European Molecular Biology Laboratory, Meiserhofstr. 1, Heidelberg,
 69117, GERMANY

Revised by author 20-JUL-2001
 Location/Qualifiers

1..20435
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1q42.1"
 /cell_type="cardiomyocyte"
 /tissue_type="cardiac"
 /dev_stage="adult"
 1..20435
 /gene="OBSCN"
 72..19934
 /gene="OBSCN"
 /function="titin binding"
 /codon_start=1
 /evidence=experimental
 /product="obscurin"
 /protein_id="CAC44768.1"
 /db_xref="GI:15026974"
 /db_xref="GOA:Q96AA2"
 /translation="MDQPOFSGAPRFLTRPKAFVSVGKDATLSCOIVGNPTPOVSWE
 KDQPVTAGARFRLAODGLYRLTLDLALGDSGOYVCRARNAIGFAFAVGLVDAAE
 ACAEQAPELRLPSTIRVREGSEATFCRVGSGPRPVSWSKDRRLSEPGPRVRY
 EELGEASALIRARAFRDGTIEVRAENPLGAASAAALVDSDAADTASRPGTSTAA
 LLAHQRRRRAMRAEGAPSPSTGRTCTVTEGKHARLSCVYTGPKPETVMKDGQ
 LVTEGRRHVYEDAQENFKILFCQSDRGLYFTASNLVGLTQTSVSVLVVREPAV
 FKRLQDLEVREKESATFLCEVPQSPTEAAWFKETRLWASAKYGIIEEGTERRLLTVR
 NVSADDDAVYICETPEGSTVAELAQGNLLRLPKRTAVRVDGTAMFCVELAVPVGP

VHLNRQEEVAVAGVVAISAEGRHTLTISQCCLEDDVGOVAFVAGDCOTSTRFCVSA
 RKPELOPDPVDPVAKARMSVILSWSPPHGERPVTIDGVLVEKKLGTITWIRCHEA
 EWATPELTADVAVAEENGFQFVSALNSGOSPYLPEPCTVHLAKPLAVRPLKAVQA
 VEGEVTFVLDLTVASAGWFLDQALKASSVYIEHCDTRHTLTITVREPASLHGQAL
 KFWANGIESSIRMEVRAAPGLTANKPAAAREVRLARLHEEAQOLLAELSQQAAATWL
 KDMGTSPGKYVQASARRVLLVRDVARDDAGLYECVSGRGIAYOISVQGLARFL
 HKDMGASDVAAGPAGFCETSEAHVHWYKDGWELGHSGERFLOEDVGTGRHLV
 AATVTRDEGTQSCRVEDSPRLRVSEPKVFAKEQLARRKLOAEAGASATLSCEV
 AQAQTEVTVKDGKLLSSSKVMEATGCTFRLVVOQADADAGEYSCAGGQRLSFV
 LDYKEPKVPAKQVAVHSEVOAEGANATLCEVAQAQAEVHWYKDGKLLSSLVHV
 BAKGCHRLVVOQAGKTDAGYCEAGQQRVSRFLHTTPEKMFKEQKLSVHVEQBA
 GASALMCEVAQAGTEVTVKDGKLLSSSKVMEVGTCTRLVLVFOAGADAGEYSC
 BAGQVRVFLHTTEPGVFAKQSVHNEVOAEGATMLCEVAQFQTEVTVKDGK
 KLSSSKVRMEVGTCTRLVVOQKADAGEYCEAGQQRVSRFLHTTPEKMFKEQ
 LVHNEVTEAGASATLCEVAQAGTEVTVKDGKLLSSSKVRIEAGQRLVVOQA
 GQADAGYCEAGQQRVFLDYSEKAVFAKQELARVQVAAAGATLSCEVAQAQ
 TEVTVKDGKLLSSSKVMEVGTCTRLVVOQADADAGEYSCAGGQRLSFV
 EPKVFAKEQVHREVOAAGASATLCEVAQAGTEVTVKDGKLLSSLVHV
 CTRLVVOQADADAGEYCEAGQQRVSRFLHTTPEKMFKEQKLSVHVEQBA
 TLCEVAQAGTEVTVKDGKLLSSSKVMEVGTCTRLVVOQADADAGEYSCAGG
 QRLSFVHVEQBAQVFAKEQVHREVOAAGASATLCEVAQAGTEVTVKDGKLLSS
 SKVRVFAKEQVHREVOAAGASATLCEVAQAGTEVTVKDGKLLSSLVHV
 PLVKEHEDIIITATLATPSAATVWLKDGVEIRRSKRHETASQGTHTLVHGAOVL
 DSATYSCVAGAGQDFVQVEEVAAKFCRLLEPVCGELGTVTLACELAPACAEVVR
 CQNTQPRVGRFQVAGRPVRSITVLGLRADAGEYCEAGGQRLSFVHVEQBA
 FMGLSTVVAEGBEITQCVVSPSDVAVWFRDGLLQSEKFAISQSGASHLTS
 DLVLEDAQITVEAGASSAALRVREAPVLFKKLEPTEPVESSRSVLEVEITPWP
 ELWRNATLAPAKNYEIIHAEGARHLVLMVGFADPFGFCTETDDTKQAKLTVM
 ROVLVRVAGCTRLVVOQADADAGEYCEAGQQRVSRFLHTTPEKMFKEQ
 PLVKEHEDIIITATLATPSAATVWLKDGVEIRRSKRHETASQGTHTLVHGAOVL
 LTLGLRPEQSGVLMVFAEGVHTSARLVVTELPSVSRQDVTVTEKEKVTCEBLS
 RPNVDVWLKDGVELRAGTMAAQAQACHSLIYCEFAQDVGVVCDADAOSSASV
 KVQRTYTLIRVLAADAGEIQVVAENASRAQLRVKELPVLTVPLRDKKLMEKHR
 GVLEQVSRASQVAFWFGQELQPKYELSDGLYKLIISDVAEDEDVTVCDAG
 DVKTSQFVVEQGITVIRGLQDVTVMEPAPMPECETSPSVRRPKWKLTVLQAG
 GNVLEQGETVRLMLRRTCTMTGTVHFTVTKGSRSSARLVSDIIPVLTRPLEPKTG
 REQSVLSLDCFPAPKAVQVYKDDTLPLQADAEAGWASFCSELSHEDVEWELNMGV
 RCOAGSHASTEVTVAREVTVTGLODAEATEGWAESFSELSHEDVEWELNMGV
 LYNDSPHEISHKGRHTLVLSIQRADAGIVRASLLKYSTASRLVYKVEVPLKALD
 DLSAEGRVSEETPARVVRHDLHVGTIKRLKTMVELEGESCFEVLSHESADPAMW
 AGLTVHVGSEETPARVVRHDLHVGTIKRLKTMVELEGESCFEVLSHESADPAMW
 TVGKTVQSSRSFOATQGRKYILVREAAAPSDAGEYVFSVGRILSKASLIVERPAA
 IIKPLEQVWAPGDEBLRCELRSAGTFVHLKDRKAIKRSQYDVYCEGTAMLVIR
 GASLKAGEYCEVASKSTASLHVEKANCFTBELTNLOVEEKAGTAVFCCKTEHPAA
 TVWKGKLLRLESKHQPQOEGTLRLTISALEKADSDTYTCDIGOAQSRALLVQV
 RVVHIIDLEDVDVQEGSSATFRCRISPNAYEPVHVELDKTPLHANELNEIDAQGY
 HVLTQLALKDSQITTFEAGDQASALRVTEKPSVFSRELDTALITEGEDTLVCE
 TSTCDIPMCVTKDGLTUGSARCOLSHEGHAQLITGATLQDSQYKNNVLRFGKYSURQ
 IVRVHAPRVFQELXDLVLEGGATLRCVLSVAAPVKWYCNVYLRFGKYSURQ
 EGAMLELVRLRLPQDSGRYSCFDQTTSATLTVTALPAQFIKLNKEATEGATAT
 LRCELSTAPVWRKGETLDRDGYCLRODQAMCELOIRGLAMVDAEYSCVCGESL
 TSASITRPMFAHFIGLRHOESLEGATATLRCELSKAAPEVWRKGRSLRDGHSR
 RQDGAVALQICGLADADAGEYSCVCGEERTSATLTVKALPAKFTGLRNEAEVAGT
 AMWLCSLAPVWRKGPENLRDGYTLRQEGTRCELQICGLAMADAGEYLCVCGQ
 ERTSATITRALPARFIEDVKNQAREGATVIOCELSNAPVWRKGPENLRDGYTLRQ
 SLRODGTCELQIRGLAMADAGEYSCVCGEERTSATLTVKALPAKFTGLRNEAEVAGT
 ATAVLRCELSKAAPEVWRKGETLDRDGYCLRODQAMCELOIRGLAMVDAEYLCW
 KERTSMLTVRAMPKSFIEGLRNEATEGATATLRCELSKAAPEVWRKGETLDRDGY
 RSHLRQDSRCELQIRGLAMADAGEYSCVCGEERTSATLTVKALPAKFTGLRNEAEVAGT
 EGATAVLQCELSKAAPEVWRKGETLDRDGYCLRODQAMCELOIRGLAMVDAEYLCW
 VCGERTSATITRALPARFIEDVKNQAREGATVIOCELSNAPVWRKGPENLRDGYTLRQ
 ANGREPLOSCTALVLDQIREDTEYTCGQSATSATLTVKALPAKFTGLRNEAEVAGT
 SVRODGTCELQIRGLAMADAGEYSCVCGEERTSATLTVKALPAKFTGLRNEAEVAGT
 BVDEGTHLCELSRAGASVWRKGETLDRDGYCLRODQAMCELOIRGLAMVDAEYLCW
 DYTCDGTQMSALSVVRPFKTRFLOESLEGATATLRCELSKAAPEVWRKGETLDRDGY
 EGVELHAGKPMRQSGATRELLIHOEAQKDTGAYCVTGQKTAASLRVTEPVTI
 RGLVDAEYVTEADYEDFCEVSRAGTVQWCLQGLPLQSNVEVTEVADGRVIRHLK
 GYTPEDAGTVFHLGNHASSQLTVRAPEVITLPLEQVQLSEQDASFCQRLSRAG
 QEARWALGGVPLQANEMNDITVEQTLHLTLHLKVTLEDAGTVSFVGTCSBAQLKV
 TAKNTVRLGNLENALEGEALFECOLSQPEVAATWLLDDEPVRTSENAEYVFFENG
 LRHLLLKRLPQDSRVTVLAGDQVTSFATLTVRGRLEILEPLKNAARVAGAQART
 CTLSAEVAVGASVYINGAAVQPDSDWTVTADGSHOALLRLSQAHPHAGEVTFACRD
 AVASHLTVLGLPUPPEDAEVAVHSHVTLMSWAAVPMSDGGGGLGCVKVEGATCQ
 WLSARLTVPGCEVDCIARGETVFRVAAVGPVAGEPVHLQVTRLAPRPVPPQ
 FSAPESRQVAAEGVSLVLEVVAAEGEVHKGMRVQPGRFVSVSQGQQLVING

ORIGIN

| | | | | | | | |
|-----------------------|--------|---|---------|------------|---|--------|-------|
| Query Match | 100.0% | Score | 385 | DB | 9 | Length | 20435 |
| Best Local Similarity | 100.0% | Pred. No. | 3.3e-78 | | | | |
| Matches | 385 | Conservative | 0 | Mismatches | 0 | Indels | 0 |
| Gaps | 0 | | | | | | |
| QY | 1 | ACATAGCCTGAGCGAGGACGGGGCTGTGTGCGAGCTGCAGATCTGTGGCTGGCTGGCTGGC | 60 | | | | |
| DB | 10286 | ACATAGCCTGAGCGAGGACGGGGCTGTGTGCGAGCTGCAGATCTGTGGCTGGCTGGCTGGC | 10345 | | | | |
| QY | 61 | AGATGCTGGGGAGTACTCTCTGTGTGTGGGAGGAGAGGACCTCTGCCACTCTCACCGT | 120 | | | | |
| DB | 10346 | AGATGCTGGGGAGTACTCTCTGTGTGTGGGAGGAGAGGACCTCTGCCACTCTCACCGT | 10405 | | | | |
| QY | 121 | GAAGGCCCTCCAGCAAGATTACAGAGAGGTCTGAGGAATGAAGAGCGCTGGAAGGGGC | 180 | | | | |
| DB | 10406 | GAAGGCCCTCCAGCAAGATTACAGAGAGGTCTGAGGAATGAAGAGCGCTGGAAGGGGC | 10465 | | | | |
| QY | 181 | CACAGCCATGTTGTGTGTGAACCTGAGCAAGTGGCCCTGTGGAGTGGAGGAAGGGCC | 240 | | | | |
| DB | 10466 | CACAGCCATGTTGTGTGTGAACCTGAGCAAGTGGCCCTGTGGAGTGGAGGAAGGGCC | 10525 | | | | |
| QY | 241 | CGAGAACCTCAGAGATGGGGACAGATACATCTCTGAGGAGGAGGAGGACAGGTGTGAGCT | 300 | | | | |
| DB | 10526 | CGAGAACCTCAGAGATGGGGACAGATACATCTCTGAGGAGGAGGAGGACAGGTGTGAGCT | 10585 | | | | |
| QY | 301 | CGAGATCTGTGGCTGGCCATGGCGAGCCCGGGAGTACTTGTGTGTGCGGCGAGCA | 360 | | | | |
| DB | 10586 | CGAGATCTGTGGCTGGCCATGGCGAGCCCGGGAGTACTTGTGTGTGCGGCGAGCA | 10645 | | | | |
| QY | 361 | GAGGACCTCAGCCACGCTCACCATC | 385 | | | | |
| DB | 10646 | GAGGACCTCAGCCACGCTCACCATC | 10670 | | | | |

RESULT 3
 HSA314905
 LOCUS
 DEFINITION

HSA314905 34667 bp DNA linear PRI 21-MAY-2002
 Homo sapiens partial OBSCN gene for obscurin, exons 31-56.

ACCESSION AJ314905
 VERSION AJ314905.1 GI:21104337
 KEYWORDS OBSCN gene; obscurin.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Young, P., Ehler, E. and Gautel, M.
 TITLE Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 34667)
 AUTHORS Gautel, M.S.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie, Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Strasse 11, Dortmund, 44227, GERMANY
 FEATURES
 source
 1. .34667
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 247..34433
 /gene="OBSCN"
 join(247..519,1332..1598,1817..2082,4388..4664,5087..5353,6008..6274,6375..6638,6939..7202,10212..10475,10837..11100,11667..11930,12466..12729,13094..13357,16768..17031,17573..17836,18196..18459,21989..22252,22617..22880,23476..23700,24686..24952,25215..25481,25573..25846,26423..26697,27413..27609,27798..27873,34161..34433)
 /gene="OBSCN"
 join(<247..519,1332..1598,1817..2082,4388..4664,5087..5353,6008..6274,6375..6638,6939..7202,10212..10475,10837..11100,11667..11930,12466..12729,13094..13357,16768..17031,17573..17836,18196..18459,21989..22252,22617..22880,23476..23700,24686..24952,25215..25481,25573..25846,26423..26697,27413..27609,27798..27873,34161..34433)
 /gene="OBSCN"
 /codon_start=3
 /product="Obscurin"
 /protein_id="CAC85752.1"
 /db_xref="GI:21104338"
 /db_xref="SPTREMBL:Q8NH3"
 /translation="LVHGIIRKLTWEVLEGSCSFECVLSHESASDPAMVTVGGKTV GSSRFQATKQGRKYLIVREAPSPDAGEVVFVRGLTSKASLIVRERPAALIKPLED QWAPGDEVLRCELRAGTFPHLKDRAIKRSQKYDVVCEGTWAMLVIRGASLNDAG EYTCVEASKESTASLHVEKANCFTLELNLQVEEKGTAVFTCKTEHPAATVTRKGI LLELRASGKHQSGELTLRLTISALEKADSDTYCDIGQAQSAQLLVQAQRRVHHI EDLEDVQSGSATFCRI SPANRVRVHFLDKTPLHANELNEIDAPQGVHVLTLR QALAKDSGTLYFAGDQASAAALRVTEKTSVFSRELTDAITTEGEDLTILVCSITCDI PVCTWKDKILRSARCLSHQHEGRAQLITGATLQDSRYKCEAGACSSITRVHA RPNVRQALDKLEVGGAATLRLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLE LVNRLRPODSGYSCSFGDQTT SATLTVTALPAQFIGLRNKEATEGATLRLCELS KAAPVWRKSGSETLRDGRYCLRDGMCELOIRGLAMVDAAEYSCVCGEERTSALT IRMPHAFIRGLRHQESLEGATATLRCELS KAAPVWRKSGRESLRDGRHSLRQDQAV CELQICGLAVDAGEYSCVCGEERTSATLTVALPAKFTTEGLRNEEAVEGATMANCE LSKVAPVWRKSGPENLRDGRYILRQEGTRCELOICGLAMADAGEYLCVCGERTSAT LTTRALPARFIEBVKQAEAREGATVLCQELNSAAPVWRKSGSETLRDGRYSLRQDG TKCELOIRGLAMADAGEYSCVCGEERTSATLTVALPDKFTTEGLRNEEATEGATVLR CELSKAPVWAMWGHETLRDGRHSLRQDQARCELOIRGLVAEDAGEYLCVCGERTS AMLTVRAMPKSFTEGLRNEEATEGDTATMLCELSKAAPVWRKSGHETLRDGRHSLRQ DGRCELQIRGLAVDAGEYSCVCGEERTSATLTVALPARFIEDVKNQAEAREGATV LQCELSKAAPVWRKSGSETLRDGRYSLRQDGRCELOIHTGLSVADTGEYSCVCGEER TSATLTVALPARFTQDLTKAEASEGATATLQCELSKAAPVWRKSGPETLRDGRYSL KQDGTCELOIHSLVDAGEYSCVCGEERTSATLTVALPARFTEGLRNEEAMEGAT ATLQCELSKAAPVWRKSGLEALRDGRKYSILRQDQAVCELOIHTGLAMADNGYSCVCGQ ERTSATLTVALPARFIEDVKNQATEGATVTLQCKRKAAPVWRKSGENTLKDGDRY SLRQDGTSCLOIRGLVIADAGEYSCICEERTSATLTVALPARFIEDVRNHEATG ATAVLQCELSKAAPVWRKSGSETLRDGRYSLRQDGRCELOIRGLVAEDTGEYLCVC QGERTSATLTVALPARFIDNMTNQAEAREGATATLHCELSKAAPVWRKSGPETLRDGD

RHSLRQDGTTRCELOIRGLSVADAGEYSCVCGEERTSATLTIREATEGATAMLQCELSK VAPVWRKSGPETLRDGRYSLRQDGRCELOIHTGLSVADTGEYSCVCGEERTSATLTV KAPOVFPREPLQSLQAEBSGSTATLOCELSSEPTATVWMSKGLQLOANGRRERPLQGGT AELVLQDLQREDTGEYTCCTCGSATSATLTVTAPVPRFLRELQHQVEDEGTAHLCCCE LSRAGASVWRKSGSLQFPFCARYQNVQDGAALLLVRGVEQEDAGDYTCOTGHTQSMAS LSVVRPFRFKTRLRQSLQESLEGTDIARLCCQLSDAESGAVVQMLKGEVTLRGLVDAEYTEM RSQATRELLIHQLEAKDTGEYACVTCQCKTAASLRVTGTEPEVTIIRGLVDAEYTADE DVEFSCVSRAGATGVQWLOGLPLOSNEVTEVAVRDGRIHTLRKQVTPEDAGTVSF HLGNHASAOALTVRAPVETILEPLQDVOLSPQDASFOCRISLRASGQAEARWALGGVPL CANENMDITVQSGTLLHLTLHKVTLERDAGTVSFHVGTCSSQAQLKVTAKNTVVRGLENE VEALEGGALFECQLSQSEVAHAHTLLDDEPHTISENAEVVFFENGLHLLLLKNLRP QDSRVTFLAGDMVTSAPLTVR"
 247..519
 /gene="OBSCN"
 /number=31
 520..1331
 /gene="OBSCN"
 /number=31
 1332..1598
 /gene="OBSCN"
 /number=32
 1599..1816
 /gene="OBSCN"
 /number=32
 1817..2082
 /gene="OBSCN"
 /number=33
 2083..4387
 /gene="OBSCN"
 /number=33
 4388..4664
 /gene="OBSCN"
 /number=34
 4665..5086
 /gene="OBSCN"
 /number=34
 5087..5353
 /gene="OBSCN"
 /number=35
 5354..6007
 /gene="OBSCN"
 /number=35
 6008..6274
 /gene="OBSCN"
 /number=36
 6275..6374
 /gene="OBSCN"
 /number=36
 6375..6638
 /gene="OBSCN"
 /number=37
 6639..6938
 /gene="OBSCN"
 /number=37
 6939..7202
 /gene="OBSCN"
 /number=38
 7203..10211
 /gene="OBSCN"
 /number=38
 10212..10475
 /gene="OBSCN"
 /number=39
 10476..10836
 /gene="OBSCN"
 /number=39
 10837..11100
 /gene="OBSCN"
 /number=40
 11101..11666
 /gene="OBSCN"
 /number=40
 11667..11930

```

/gene="OBSCN"
/number=41
11931..12465
/gene="OBSCN"
/number=41
12466..12729
/gene="OBSCN"
/number=42
12730..13093
/gene="OBSCN"
/number=42
13094..13357
/gene="OBSCN"
/number=43
13358..16767
/gene="OBSCN"
/number=43
16768..17031
/gene="OBSCN"
/number=44
17032..17572
/gene="OBSCN"
/number=44
17573..17836
/gene="OBSCN"
/number=45
17837..18195
/gene="OBSCN"
/number=45
18196..18459
/gene="OBSCN"
/number=46
18460..21988
/gene="OBSCN"
/number=46
21989..22252
/gene="OBSCN"
/number=47
22253..22616
/gene="OBSCN"
/number=47
22617..22880
/gene="OBSCN"
/number=48
22881..23475
/gene="OBSCN"
/number=48
23476..23700
/gene="OBSCN"
/number=49
23701..24685
/gene="OBSCN"

Query Match      67.8%; Score 261; DB 9; Length 34667;
Best Local Similarity 100.0%; Pred. No. 7.3e-50;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 125  GCCCTGCCAGCAAGTTCACAGAGGCTCTGAGGAATGAAGAGCCCTGGAAGGGGCCACA 184
Db 10211 GCCCTGCCAGCAAGTTCACAGAGGCTCTGAGGAATGAAGAGCCCTGGAAGGGGCCACA 10270

Qy 185  GCCATGTTGGTGTGAACAGTGAAGTGGCCCTGTGGAGTGGAGAGGGGCCGAG 244
Db 10271 GCCATGTTGGTGTGAACAGTGAAGTGGCCCTGTGGAGTGGAGAGGGGCCGAG 10330

Qy 245  AACCTCAGAGTGGGACAGATACATCTCAGGAGGAGGGACCAAGTGTGAGCTGCAG 304
Db 10331 AACCTCAGAGTGGGACAGATACATCTCAGGAGGAGGGACCAAGTGTGAGCTGCAG 10390

Qy 305  ATCTGTGGCTGCCATGGCGGACCGGGAGTACTTGTGTGTGTCGGGGCAGGAGG 364
Db 10391 ATCTGTGGCTGCCATGGCGGACCGGGAGTACTTGTGTGTGTCGGGGCAGGAGG 10450

Qy 365  ACCTCAGCAGCGCTCACCATC 385
Db 10451 ACCTCAGCAGCGCTCACCATC 10471

RESULT 4
AL353593      135964 bp      DNA      linear      PRI 13-FEB-2002
LOCUS      Human DNA sequence from clone RP5-1139B12 on chromosome 1q42.1-43,
DEFINITION      complete sequence.
ACCESSION      AL353593
VERSION      AL353593.33 GI:18673899
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 135964)
AUTHORS      Almeida,J.
TITLE      Direct Submission
JOURNAL      Submitted (13-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonesrequest@sanger.ac.uk
On Feb 14, 2002 this sequence version replaced gi:17977879.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em; EMBL; Sw;
SWISSPROT; Tr; TrEMBL; Wp; WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP5-1139B12 is from the library RPCI-5 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pcYPAC2
This sequence is the entire insert of clone RP5-1139B12.
FEATURES
Location/Qualifiers
1..135964
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.1-43"
/clone="RP5-1139B12"
/clone_lib="RPCI-5"
25928..26054
/misc_feature
/note="Sequence from overlapping clone RP11-520H14
(AL359510). Assembly confirmed by restriction digest."

ORIGIN

Query Match      67.8%; Score 261; DB 9; Length 135964;
Best Local Similarity 100.0%; Pred. No. 5.7e-50;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 125  GCCCTGCCAGCAAGTTCACAGAGGCTCTGAGGAATGAAGAGCCCTGGAAGGGGCCACA 184
Db 79224 GCCCTGCCAGCAAGTTCACAGAGGCTCTGAGGAATGAAGAGCCCTGGAAGGGGCCACA 79283

```

```

Qy 185 GCCATGTTGTGGTGAACCTGAGCAAGGTGGCCCTCTGTGAGTGGAGGAAGGGCCCGAG 244
Db 79284 GCCATGTTGTGGTGAACCTGAGCAAGGTGGCCCTCTGTGAGTGGAGGAAGGGCCCGAG 79343
Qy 245 AACCTCAGAGATGGGACAGATACATCTCTGAGGCGAGGAGGGACCGAGTGTGAGCTGCAG 304
Db 79344 AACCTCAGAGATGGGACAGATACATCTCTGAGGCGAGGAGGGACCGAGTGTGAGCTGCAG 79403
Qy 305 ATCTGTGCTGCGCCATGCGGAGCGCCGGGAGTACTTGTGTGTGTCGGCGGACGAGAGAG 364
Db 79404 ATCTGTGCTGCGCCATGCGGAGCGCCGGGAGTACTTGTGTGTGTCGGCGGACGAGAGAG 79463
Qy 365 ACCTCAGCCAGCTCACCATC 385
Db 79464 ACCTCAGCCAGCTCACCATC 79484

RESULT 5
AC026657 164766 bp DNA linear HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 1 clone RP11-245P10, WORKING DRAFT
DEFINITION AC026657
ACCESSION AC026657
VERSION AC026657.4 GI:9958202
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
2 (bases 1 to 164766)
Unpublished
Waterston,R.H.
Direct Submission
Submitted (22-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7637349.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0245P10
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 139884 bases at least Q40
Consensus quality: 147686 bases at least Q30
Consensus quality: 151469 bases at least Q20
Insert size: 169000; agarose-fp
Insert size: 161074; sum-of-ctngs
Quality coverage: 3.60 in Q20 bases;
Quality coverage: 3.92 in Q20 bases; sum-of-ctngs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1124: contig of 1124 bp in length
* 1125 1224: gap of unknown length
* 1225 3032: contig of 1808 bp in length
* 3033 3132: gap of unknown length

FEATURES
source
1..164766
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-245P10"
1..1124
/note="assembly_name:Contig14"
1225..3032
/note="assembly_name:Contig21"
3133..4493
/note="assembly_name:Contig23"
4594..5860
/note="assembly_name:Contig24"
5961..7670

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```

```
misc_feature /note="assembly_name:Contig25"
7771..9669
/note="assembly_name:Contig26"
9770..11715
/note="assembly_name:Contig27"
11816..14243
/note="assembly_name:Contig28"
14344..16887
/note="assembly_name:Contig29"
16988..19247
/note="assembly_name:Contig30"
19348..21375
/note="assembly_name:Contig31"
21476..25025
/note="assembly_name:Contig32"
25126..28051
/note="assembly_name:Contig33"
28152..32054
/note="assembly_name:Contig34"
32155..35716
/note="assembly_name:Contig35"
35817..40891
/note="assembly_name:Contig36"
40922..44027
/note="assembly_name:Contig37"
44128..48899
/note="assembly_name:Contig38"
49000..54255
/note="assembly_name:Contig39"
54356..58994
/note="assembly_name:Contig40"
59095..64390
/note="assembly_name:Contig41"
64491..70865
/note="assembly_name:Contig42"
70966..78667
/note="assembly_name:Contig43"
78768..87207
/note="assembly_name:Contig44"
87308..97858
/note="assembly_name:Contig45"
97959..109520
/note="assembly_name:Contig46"
109621..120908
/note="assembly_name:Contig47"
121009..141477
/note="assembly_name:Contig48"
141578..162672
/note="assembly_name:Contig49"
162773..163974
/note="assembly_name:Contig9"
164075..164766
/note="assembly_name:Contig7"
```

ORIGIN

```
Query Match 67.8%; Score 261; DB 2; Length 164766;
Best Local Similarity 100.0%; Pred. No. 5.5e-50;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 GCCTGCCAGCAAGTTCACAGAGGTTCTGAGGAATGAAGAGGCCGCTGGAAGGGGCCACA 184
D5 50329 GCCTGCCAGCAAGTTCACAGAGGTTCTGAGGAATGAAGAGGCCGCTGGAAGGGGCCACA 50388
QY 185 GCCATGTTGTGTGTGTAAGTCTGAGCAAGTGGCCCTGTGAGTGGAGGAGGGGCCGAG 244
D5 50389 GCCATGTTGTGTGTGTAAGTCTGAGCAAGTGGCCCTGTGAGTGGAGGAGGGGCCGAG 50448
QY 245 AACCTCAGAGATGGGGACAGATACATCTCTGAGGAGGAGGAGGACCAAGTGTGAGCTCAG 304
D5 50449 AACCTCAGAGATGGGGACAGATACATCTCTGAGGAGGAGGAGGACCAAGTGTGAGCTCAG 50508
QY 305 ATCTGTGCGCTGGCCATGGCGACCGCGGGAGTACTTTGTGTGTGTCGGGACAGGAGG 364
```

```
Db 50509 ATCTGTGCGCTGGCCATGGCGACCGCGGGAGTACTTTGTGTGTGTCGGGACAGGAGG 50568
QY 365 ACCTCAGCAGCGCTCACCATC 385
D5 50569 ACCTCAGCAGCGCTCACCATC 50589

RESULT 6
AC023889/c 174612 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT
DEFINITION SEQUENCE, 13 unordered pieces.
AC023889
AC023889.3 GI:8969253
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174612)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 174612)
Waterston,R.H.
Direct Submission
Submitted (18-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8748947.

COMMENT
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH061B12
----- Summary Statistics -----
Sequencing vector: W13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 148000; agarose-fp
Insert size: 173412; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.55 in Q20 bases; sum-of-contigs
-----
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1
* 2478: contig of 2478 bp in length
* 2479
* 2578: gap of unknown length
* 2579
* 6948: contig of 4370 bp in length
* 6949
* 7049: gap of unknown length
* 7049
* 9373: contig of 2325 bp in length
* 9374
* 9473: gap of unknown length
* 9474
* 13303: contig of 3830 bp in length
* 13304
* 13403: gap of unknown length
* 13404
* 18270: contig of 4867 bp in length
* 18271
* 18370: gap of unknown length
* 18371
* 25589: contig of 7219 bp in length
* 25590
* 25689: gap of unknown length
* 34815: contig of 9126 bp in length
* 34816
* 34915: gap of unknown length
```


Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Apr 7, 2002 this sequence version replaced gi:19699553.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-192P17 is from the RP23-23 Mouse PAC Library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

FEATURES

source

Location/Qualifiers

1..211829
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="11"
 /clone="RP23-192P17"
 /clone_lib="RPCI-23"

ORIGIN

Query Match 50.9%; Score 196; DB 10; Length 211829;
 Best Local Similarity 80.6%; Pred. No. 4e-35;
 Matches 229; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
 QY 102 CCTGTCCACTCTACCGTGAAGCCCTCCAGCCAGTTCACAGAGGCTGTGGAATG 161
 Db 46277 CCTGTGCTTTCTGACCTCTTACCTCTTACCTCTGCCCCACAGTTCACAGAGGCTGTGGAATG 46218
 QY 162 AAGAGGCCGTGAAGGGGCCACAGCCATGTTGTGTGTAAGTTCAGCAAGTGTGCCCTCTG 221
 Db 46217 AAGAGGCCACGAGAGGCACCATGCTGAGGTGCCAGATGAGCAAGCTGTCCCTCTG 46158
 QY 222 TGGAGTGAGGAGAGGGCCCGAGAACCTTCAGAGATGGGGACAGATACATCTTGGGCGAG 281
 Db 46157 TGGAGTGAGGAGAGGGGTACAGAGACCTTCAGAGATGGGGACAGATACAGCTTGGGCGAG 46098
 QY 282 AGGGACCAAGTGTGAGCTGCAGATCTGTGCTTGGCCATGGCGGACCGCGGGAGTACT 341
 Db 46097 ATGGGGCCATGTTGTGAGCTGCAGATCTGTGCTTGGCCATGGCGGACCGCGGGAGTACT 46038
 QY 342 TGTGTGTGTGGGCGAGGAGGACCTTCAGCCACCTCACCATC 385
 Db 46037 CATGTGTGTGTGGGCGAGGAGAGCTCAGCCACACTGAGTGTC 45994

RESULT 9

AC026657/c

LOCUS

AC026657 164766 bp DNA linear HTG 01-SEP-2000
 Homo sapiens chromosome 1 clone RP11-245P10, WORKING DRAFT

DEFINITION

SEQUENCE, 31 unordered pieces.

ACCESSION

AC026657.4

VERSION

HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

Homo sapiens (human)

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 164766)

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

TITLE

COMMENT

Waterston,R.H.
 The sequence of Homo sapiens clone
 Unpublished
 2 (bases 1 to 164766)
 Waterston,R.H.
 Direct Submission
 Submitted (22-MAR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Sep 1, 2000 this sequence version replaced gi:7637349.

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

Center project name: H NH0245P10

Sequencing vector: plasmid, 0%

Chemistry: Dye-terminator Big Dye, 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 139884 bases at least Q40

Consensus quality: 147686 bases at least Q30

Consensus quality: 151469 bases at least Q20

Insert size: 169000; agarose-fp

Insert size: 161074; sum-of-contigs

Quality coverage: 3.60 in Q20 bases; agarose-fp

Quality coverage: 3.92 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 31 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1124: contig of 1124 bp in length

* 1125 1224: gap of unknown length

* 1225 3032: contig of 1808 bp in length

* 3033 3133: gap of unknown length

* 3133 4493: contig of 1361 bp in length

* 4494 4593: gap of unknown length

* 4594 5860: contig of 1267 bp in length

* 5861 5960: gap of unknown length

* 5961 7670: contig of 1710 bp in length

* 7671 7770: gap of unknown length

* 7771 9669: contig of 1899 bp in length

* 9670 11715: contig of 1946 bp in length

* 11716 11815: gap of unknown length

* 11816 14243: contig of 2428 bp in length

* 14244 14343: gap of unknown length

* 14344 16887: contig of 2544 bp in length

* 16888 16987: gap of unknown length

* 16988 19248: contig of 2260 bp in length

* 19249 19348: gap of unknown length

* 19349 21375: contig of 2028 bp in length

* 21376 21475: gap of unknown length

* 21476 25025: contig of 3550 bp in length

* 25026 25125: gap of unknown length

* 25126 28051: contig of 2926 bp in length

* 28052 28151: gap of unknown length

* 28152 32054: contig of 3903 bp in length

* 32055 32154: gap of unknown length

* 32155 35716: contig of 3562 bp in length

* 35717 35816: gap of unknown length

* 35817 40991: contig of 5075 bp in length

* 40992 44027: contig of 3036 bp in length

* 44028 44127: gap of unknown length

Query Match 49.5%; Score 190.6; DB 6; Length 2155;
 Best Local Similarity 83.1%; Pred. No. 1.5e-33;
 Matches 217; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 125 GCGCTGCGCAGCAAGTTTCACAGAGGGTCTGAGGAATGAAGAGGCGGTGGAGGGGCCACA 184
 Db 205 GCGCTGCGCAGGTTTCACAGAGGGTCTGAGGAATGAAGAGGCGGTGGAGGGGCCACA 264

Qy 185 GCGATGTTGGTGTGTAAGTACAGCAAGTGGCCCTGAGTGGAGGAGGAGGGCCGAG 244
 Db 265 GCGACACTGCAATGTGAGTGTGAGCAAGCGACGCCCTGTGAGTGGAGGAGGAGGCTTGG 324

Qy 245 AACCTTCAGAGATGGGACAGATACATCTCTGAGGAGGAGGAGGACCGAGTGTGAGCTCGAG 304
 Db 325 GCTCTCAGAGATGGGACAAATACAGCTCAGACAGCGGGCTGTGTGAGCTCGAG 384

Qy 305 ATCTGTGGCTGGCAGTGGCGAGCGCGGGAGTACTGTGTGTCGGGCGAGGAGG 364
 Db 385 ATTATGCGCTGGCTATGGCAGATACAGCGGGTGTACTCATGTGTGTGTCGGCAGGAGG 444

Qy 365 ACCTCAGCGCAGCTCACCATC 385
 Db 445 ACCTCAGCTACACTCACTGTC 465

RESULT 11

AC099089 260998 bp DNA linear HTG 10-MAY-2003
 Rattus norvegicus clone CH230-154E3, WORKING DRAFT SEQUENCE.

AC099089.5 GI:30522701

HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 260998)

REFERENCE

AUTHORS

Munry,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
 Allen,A., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,X., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorensheva,L., Loulsec,H., Lozano,R.J., Lu,X., Ma,J.,
 Mathewwar,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Margum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mathiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
 Nwaokelimeh,O., Okwunu,G., Olarpunsgoon,A., Pal,S., Parks,K.,
 Pasceriak,S., Paul,H., Perez,A., Perez,E., Pfannkuch,C.,
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Shvartsbeyn,A., Sibson,I., Sitter,C.D., Smajs,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umami,K.,
 Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wlecyk,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstein,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 260998)

Worley,K.C.

Direct Submission

Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 260998)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GGLZ

Center clone name: CH230-154E3

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 216929 bases at least Q40

Consensus quality: 219228 bases at least Q30

Consensus quality: 220845 bases at least Q20

Estimated insert size: 226118; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 1 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 260998: contig of 260998 bp in length.

Location/Qualifiers

1..260998

/organism="Rattus norvegicus"

FEATURES

source

detection was carried out by SSAHA-SNP. 225,000 reads were annotated

as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

FEATURES

source
1..621
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="CZECHII/Ei"
/db_xref="taxon:10090"
/map="11 22-567 59585524-59586069"
/clone_lib="CZECHII/Ei"
<1..>621

STS

ORIGIN

Query Match 42.9%; Score 165; DB 11; Length 621;
Best Local Similarity 77.0%; Pred. No. 1.4e-27;
Matches 201; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 125 GCCTGCCAGCCAGTTCACAGAGGTTCTGAGGAATGAGAGGCCGTGGGAAGGGGCCACA 184
Db GCCTGCCAGCCAGTTCACAGAGGTTCTGAGGAATGAGAGGCCGTGGGAAGGGGCCACA 418
QY 185 GCCATGTTGGTGTGAACAGCAAGGTGCCCTGTGGAGTGGAGGAAGGGGCCCGAG 244
Db GCCATGTTGGTGTGAACAGCAAGGTGCCCTGTGGAGTGGAGGAAGGGGCCCGAG 358
QY 245 AACCTCAGAGATGGGACAGATATCATCTGAGGAGGAGGGGACACAGGTGTGAGCTCGAG 304
Db AACCTCAGAGATGGGACAGATATCATCTGAGGAGGAGGGGACACAGGTGTGAGCTCGAG 298
QY 305 ATCTGTGCGCTGGCCATCGCGGCGGAGTACTGTGTGTGTCGGCGCAGGAGG 364
Db ATCTGTGCGCTGGCCATCGCGGCGGAGTACTGTGTGTGTCGGCGCAGGAGG 238
QY 365 ACCTCAGCCAGCTCACCATC 385
Db ACCTCAGCCAGCTCAGGTGTC 217

RESULT 14

AC139630
LOCUS 98613 bp DNA linear HTG 08-MAR-2003
DEFINITION Takifugu rubripes clone 221D8, WORKING DRAFT SEQUENCE, 6 ordered pieces.

ACCESSION

AC139630 GI:28882132

VERSION

HTG; HIGS_PHASE2; HIGS_DRAFT.

KEYWORDS

Takifugu rubripes (Fugu rubripes)

SOURCE

Takifugu rubripes

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Takifugu.
1 (bases 1 to 98613)
Akter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J., Hachichi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laic,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B., Magulies,E.H., Mastello,C., Masker,B., McDowell,J., Paquirigan,C., Pearson,K., Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative

REFERENCE

AUTHORS

Unpublished

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

2 (bases 1 to 98613)
Direct Submission
Submitted (07-FEB-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

3 (bases 1 to 98613)
Green,E.D.
Direct Submission
Submitted (08-MAR-2003) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Mar 8, 2003 this sequence version replaced gi:28269380.

COMMENT

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: http://www.nisc.nih.gov

Contact: nisc.zoo@nih.gov

----- Project Information

Center Project name: egw

Center Clone name: 221D08

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 97841 bases at least Q40
Consensus quality: 98004 bases at least Q30
Consensus quality: 98090 bases at least Q20
Insert size: 85000; agarose-fp
Insert size: 98113; sum-of-contigs
Quality coverage: 11.51x in Q20 bases; agarose-fp
Quality coverage: 9.97x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 28467: contig of 28467 bp in length

* 28468 28567: gap of unknown length

* 28568 31135: contig of 2568 bp in length

* 31136 31235: gap of unknown length

* 31236 41487: contig of 10252 bp in length

* 41488 41587: gap of unknown length

* 41588 74365: contig of 32778 bp in length

* 74366 74465: gap of unknown length

* 74466 96541: contig of 22076 bp in length

* 96542 96641: gap of unknown length

* 96642 98613: contig of 1972 bp in length.

FEATURES

source

1..98613
/organism="Takifugu rubripes"
/mol_type="genomic DNA"
/db_xref="taxon:31033"
/clone="221D8"
/clone_lib="Incyte Genomics"

misc_feature

1..28467
/note="assembly_fragment"

vector_end:77

vector_side:left

misc_feature

28568..31135
/note="assembly_fragment"

misc_feature

31236..41487

```

/note="assembly_fragment"
41588..74365
/note="assembly_fragment"
74466..96541
/note="assembly_fragment"
96642..98613
/note="assembly_fragment
clone end:SP6
vector_side:right"

ORIGIN
Query Match      18.5%; Score 71.4; DB 2; Length 98613;
Best Local Similarity 57.0%; Pred. No. 1.5e-06;
Matches 151; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

QY 120 TGAAGGCCCTCCAGCCAGTTTCACAGAGGCTTCAGGAATGAAGAGGCCGCTGGAAGGGG 179
DB 61703 TGACAGGCTTACCAACCGACTTTCAGACGGCTCTCATCAACAGGAAGTGTATGAGGGCA 61762

QY 180 CCACAGCCATCTTGTGTGTGAACCTGAGCAAGTGGGCCCT---GTGGAGTGGAGGAAGG 236
DB 61763 ACAGCGTATTCTGCACCTGTGAGCTCAGCAAGCTGTTCTTCGGTGGAGTGGAGGAGAG 61822

QY 237 GGGCCGAGAACCTCAGAGATGGGACAGATACATCCTGAGGCGAGAGGGACACAGGTGTG 296
DB 61823 GAGGAGAGCTCTAAAGCATGGAGACAAATATCAGGTGAGGAAAAAAGAGCTGCAGGTGG 61882

QY 297 AGCTGCAGATCTGTGGCTGGCCATGGCGGACGCCGGGAGTACTTGTGTGTGTGGGGC 356
DB 61883 AGATGAAGATCTCGACCTCAGTCTCAGCGACGCGGAGAGTACAGCTGTGTGTGGAG 61942

QY 357 AGGAGAGGACCTCAGCCACGCTCAC 381
DB 61943 AACAGGCAACGAGAGCGCGTGATCAC 61967

RESULT 15
AX834822 LOCUS AX834822 2534 bp DNA linear PAT 15-DEC-2003
DEFINITION Sequence 1946 from Patent EP1347046.
ACCESSION AX834822
VERSION AX834822.1 GI:39920957
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
Yanamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Naganari, K. and
Masuho, Y.
TITLE Full-length cDNA sequences
JOURNAL Patent: EP 1347046-A 1946 24-SEP-2003;
Research Association for Biotechnology (JP)
FEATURES
Source 1..2534
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match      18.2%; Score 70; DB 6; Length 2534;
Best Local Similarity 51.4%; Pred. No. 6e-06;
Matches 188; Conservative 0; Mismatches 175; Indels 3; Gaps 1;

QY 17 GACGGGCTGTGCGGAGCTGCAGATCTGTGGCTGCTGTGCAGATGTGGGAGTAC 76
DB 937 GAGGGCACGATGGCCATGCTGTGTATCCGGGGGCCCTCGCTCAAGGACGCGGGCGAGTAC 996

QY 77 TCCGTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCGTGAAGGCCCTGCCAGCC 136
DB 997 ACGTGTGAGGTGGAGGCTTCCAGAGGACAGCCAGCCCTCCATGTGGAGAAAAGCAAC 1056

```

Search completed: September 19, 2004, 19:27:07
Job time : 1574.62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 14:16:08 ; Search time 4363.43 Seconds
(without alignments)
10628.572 Million cell updates/sec

Title: US-10-077-130-4_COPY_1_1070
Perfect score: 1070
Sequence: 1 tgcctaccagcagccacac.....tagtgcgagccgcgggt 1070

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

. GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_nu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|-----------|-------------|
| 1 | 1070 | 100.0 | 20435 | 9 | HSJ2535 | Homo sapi |
| 2 | 1005 | 93.9 | 95745 | 9 | AL359510 | Human DNA |
| 3 | 720 | 67.3 | 211829 | 10 | AL645854 | Mouse DNA |
| 4 | 699.2 | 65.3 | 260998 | 2 | AC099089 | Rattus no |
| 5 | 527.6 | 49.3 | 44358 | 2 | AC142478 | Rattus no |
| 6 | 505 | 47.2 | 7531 | 9 | HSJ314896 | Homo sapi |
| 7 | 291.4 | 27.2 | 1198 | 11 | WU409657 | Mus muscu |
| 8 | 255 | 23.8 | 2884 | 6 | AX714317 | Sequence |
| 9 | 255 | 23.8 | 2884 | 9 | AK056556 | Homo sapi |
| 10 | 199.2 | 18.6 | 3316 | 9 | BC007201 | Homo sapi |
| 11 | 197.8 | 18.5 | 135033 | 9 | AC009955 | Homo sapi |
| 12 | 197.8 | 18.5 | 186883 | 2 | AC040991 | Homo sapi |
| 13 | 181 | 16.9 | 225471 | 2 | AC112361 | Rattus no |
| 14 | 181 | 16.9 | 266167 | 2 | AC121633 | Rattus no |
| 15 | 166.8 | 15.6 | 187999 | 2 | AC115011 | Mus muscu |
| 16 | 158.6 | 14.8 | 2141 | 10 | S56581 | alpha inhib |
| 17 | 126 | 11.8 | 5382 | 9 | AB014557 | Homo sapi |
| 18 | 117.2 | 11.0 | 471 | 6 | BD060356 | Secreted |
| 19 | 108.8 | 10.2 | 153945 | 2 | BX322540 | Danio rer |
| 20 | 99.6 | 9.3 | 82746 | 1 | AF453501 | Actinosyn |
| 21 | 98 | 9.2 | 135964 | 9 | AL353593 | Human DNA |
| 22 | 95.6 | 8.9 | 745 | 9 | HSJ33599 | Homo sapi |
| 23 | 90.6 | 8.5 | 1393 | 11 | PM11H12G | Penicilli |
| 24 | 88.4 | 8.3 | 154890 | 2 | BX548249 | Penicilli |
| 25 | 84.8 | 7.9 | 1065 | 11 | PM2B12B | Penicilli |
| 26 | 84.6 | 7.9 | 1393 | 11 | PM11H12G | Penicilli |
| 27 | 84.4 | 7.9 | 1279 | 11 | PM2H12G | Penicilli |
| 28 | 83.6 | 7.8 | 95209 | 2 | AP004323 | Oryza sat |
| 29 | 82.2 | 7.7 | 181850 | 2 | BX276102 | Danio rer |
| 30 | 81 | 7.6 | 956 | 11 | PM2D12B | Penicilli |
| 31 | 79.8 | 7.5 | 976 | 11 | PM12B6G | Penicilli |
| 32 | 79.8 | 7.5 | 1065 | 11 | PM2B12B | Penicilli |
| 33 | 79.6 | 7.4 | 956 | 11 | PM2D12B | Penicilli |
| 34 | 79.4 | 7.4 | 733 | 9 | HSJ335799 | Homo sapi |
| 35 | 79.2 | 7.4 | 301925 | 1 | AP005046 | Streptomy |
| 36 | 79 | 7.4 | 125020 | 9 | AF429315 | Homo sapi |
| 37 | 78.6 | 7.3 | 10444 | 14 | PVI422133 | Pseudorab |
| 38 | 78.4 | 7.3 | 224777 | 2 | AC138109 | Mus muscu |
| 39 | 78.2 | 7.3 | 11059 | 14 | PRVULGNS | Pseudorab |
| 40 | 77.8 | 7.3 | 869 | 11 | PM2A12B | Penicilli |
| 41 | 77.4 | 7.2 | 1052 | 11 | PM2H12B | Penicilli |
| 42 | 77.4 | 7.2 | 162832 | 5 | BX005391 | Zebrafish |
| 43 | 77.4 | 7.2 | 224777 | 2 | AC138109 | Mus muscu |
| 44 | 77.2 | 7.2 | 991 | 11 | PM12H12B | Penicilli |
| 45 | 77.2 | 7.2 | 1143 | 11 | PM12A11G | Penicilli |

ALIGNMENTS

| | | | | | |
|------------|--|-------------|------|--------|-----------------|
| RESULT 1 | HSJ2535 | 20435 bp | mRNA | linear | PRI 14-SEP-2001 |
| LOCUS | Homo sapiens mRNA for obscurin (OBSCN gene). | | | | |
| DEFINITION | Obscurin | | | | |
| ACCESSION | U002535 | | | | |
| VERSION | U002535.1 | GI:15026973 | | | |
| KEYWORDS | OBSCN gene; obscurin. | | | | |
| SOURCE | Homo sapiens (human) | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | 1 | | | | |
| AUTHORS | Young, P., Ehler, E. and Gaucel, M. | | | | |
| TITLE | Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly | | | | |

J. Cell Biol. 154 (1), 123-136 (2001)
 21342081
 MEDLINE
 PUBMED
 2 (bases 1 to 20435)
 11448995
 REFERENCE
 AUTHORS
 Gautel, M.S.
 TITLE
 Direct Submission
 Submitted (31-OCT-1997) Gautel M.S., Structural Biology Division,
 European Molecular Biology Laboratory, Meyerhofstr. 1, Heidelberg,
 69117, GERMANY
 Revised by author 20-JUL-2001
 REMARK
 FEATURES
 Location/Qualifiers
 1. 20435
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1q42.1"
 /cell_type="cardiomyocyte"
 /tissue_type="cardiac"
 /dev_stage="adult"
 1. 20435
 /gene="OBSCN"
 72. 19934
 /gene="OBSCN"
 /function="titin binding"
 /codon_start=1
 /evidence=experimental
 /product="obscurin"
 /protein_id="CAC44768.1"
 /db_xref="GI:15026974"
 /db_xref="GOA:Q96AA2"
 /db_xref="SPTREMBL:Q96AA2"
 /translation="MDQPSGAPRFLTRPKAFVSVSGKDATLSQIVGNPTQVSWSE
 KDQPVYTAGARFLAQDGLYRLTILDALGDSQVYVCARNAIGFAFVAGVLDVDAE
 AACAEQAPHELLPRTSRVREGSEATFRCHVGSGSPAPVSWKDRGLPEQGPVRV
 ELGEASALRIARPDGTYEVRAENPLGASAAALVDSADADTSRGTSTAA
 LLAHQRRAMRAEPAPSPSTGRTCTCTCTGKRAUSVDSGPKPTVTKDQG
 LTEGRHHVYDAQENFLKILFCQSDRLGTYCTASNLGGTYSSVLVVRREPAPV
 FKRLQDLVREKESATFLCEVPQSTEARAFKEETRLWASAKYGIIEEGTERTLTVR
 NVSADDDVICTPEGSTVAELAVQGNLLRLPKRTAVRGDTAMFCVELAVPVG
 VHMLNDEEVVAGGRVAISAEGRHTLTISQCCLEDVGVGVAMAGCQSTSRFCVSP
 RKPLPDPVDPVKARMSVILSWSPPHGERPPTIDGVLEKXKLTGYTRFCHEA
 EWATPELTVDVAENGNFVRNALNSFGQSPYLEPFGTVLHAPKLAVRPLKAVQA
 VEGEVFTVLDLTVASAGWFLDQALKASVYIEHCDRTHTLTIREVPASLHQAQ
 KFGVANGTESIRMEVRAAPGLTANKPAAAREVTLARLHEEAQGLTAELEDQAAATVL
 KDGLTSPGPKYEVQASGRVLLVRDVARDDAGLYECVSRGRTAYQLSVQGLARFL
 HKMAGSCDVAAGGPAQFCPTSEAHVHVHVKDGMELCHSGERFLOQDVGTRHLV
 AATVTRDEGTYSRCVGEDSDVDFRLVSRPKVVFVFAKEQLARRKLQAEASATLSCEV
 AQTQETVWKDGGKLSSSKVCMEATGCTRLVAVQAGQADAGEYSCAGGRLSPH
 LDYKEPVVFAKQVHAHSEVQAEAGANATLSCEVAQAQAEVWYKDKKLSGLKHV
 EAGKRRLLVQQAQGTADGYSCARGQKVSFLRHITBPKMMFAKEQSVHNEVQAE
 GAGAMLSCEVAQAQETVWTKDGGKLSSSKVGMEVKGCTRLVLPQAGKADAGEYC
 EAGQRVSFHLHITPEKGVFAKQSVHNEVQAEAGTAMLSCEVAQAQETVWTKDGG
 KLSSSKVRMEVKGCTRLVVOQVQADAGEYSCAGGQVSPQLHITPEKAVFAKEQ
 LVHNEVTEAGSATLSCEVAQAQETVWTKDGGKLSSSKVRIBAAQGMOLVVOQA
 QDADAGYTCAGGQRLSFLHLDVSEPKAVFAKEQLARRKVOAEAGIATLSCEVAQAQ
 TETVWTKDGGKLSSSKVRMEAVGCTRLVVOQAQADAGEYSCAGGQRLSFLDVA
 EPKVFAKEQVHREVOQAQASTLSCEVAQAQETVWYKDKGKLSKVRMEAVG
 CTRRLVVOQAQADAGEYSCAGGQRLSFLHVAEPKAVFAKEQPAQSERVQAEATSA
 TLSCEVAQAQETVWTKDGGKLSSSKVRMEAVGCTRLVVOQAQADAGEYSCAGD
 LRLSFLHVAEPKAVFAKEQPAHREVOQAQASTLSCEVAQAQETVWTKDGGKLS
 SKVRVAVGCTRLVVOQAQADAGEYSCAGGQRLSFLHVAEPKAVFAKEQVSPQCRRE
 QVLSGLRPEDSGMLVFAKEGVTYSARLVLTLPVPSRPLQDVGTVCKEADHQAQSAV
 RNVDPVRLKDGVLEBAGTMAIAAQAGCRSLTITYCEPADQGVVCEADHQAQSAV
 KVQGRYTILYRVLADAGEIQFVAENAESRAQLRVKELPTVLRPLDKLAMEKHR
 GVLEQCVRASASAVRVKGSQELQPGPKYELVSDGGLYRKLLTISDVHAEDDTYTCDAE

DVKTSAQFFVEEQSITIVRGLQDVTVMPEPAPAFECETSIPSVRPPKWLKTKTVLQAG
 GNVGLEQGTVHRLMLRRTCTMTGPHVHTVVKRSRSARLVSDIPVLTPLPLPKTG
 RELQSVLSCDFRPAKAVOVYKODTFLPSPEKFKMSLEGOMAEHLRLRLMDAGYV
 RCQAGSHSSTEVTVAREVTVTGLQDAEATEEGWASFSCELSDEDEEVMWSLNGVP
 LYNSDFHEISHKRRHTLVLSKSIQVADAGIVRASLSKVSTARSLEVRVVKPVVYKALD
 DLSAEERTLALQCEVSDPEAHVVRKDGVLGSPDKYDPLHTAGTGLVVDVSPD
 AGLTVCHVSEETARVVRHDLHVITKRLKTMELVEGSCSECVLSHESASDPAW
 TVGKTVGSSRFQATQGRKIILVREAPSDAGEVVFVRGLTNSKSLIVRRPAA
 IIKLEDQWAPADVELRCELSRAGTPVHMLKDKRKAKIRKSKYQKVCBGMVAVLVR
 CASLKDAGEYTCVEASKSTASLHEEAKNCFBELTNLQVEEKTQVAFCTKEHPAA
 TVTWKGLLELRASGKHPQSEGLTILITSALEKADSDTVCTIGQAQRAQLVQAG
 RVHIIIDLEDVDVQEGSSATFCRIIPANVEPVHFLDKTPLHNLINELIDAOPGY
 HVLRLALAKDSGTIVFEAGDORASALVRTEKPSVFSRELTDATITEGEDTLVCE
 TSTCDIPMCWTGDKTLRGSACQLSHEGHAOLITGATLQDSGRVYKCSAGGACSS
 IIVRHARVFEAKOLEVLEGGANILRCELSVVAAPVWKCIGNVVRFGDKISLRQ
 EGAMLEVVRLNRPODSGRYSCFGDQTTATLTVTALPAQFICKLNKKEGATEGAT
 LRCELSKTAPVWRKSETLRDGRYCLRDQGANCELOIRGLAVDAAEYSCVGEER
 TSASLTIRMPAHFIRGLRHOESIEGATLRLCELSKAAPVWRKGRSRLDGRHSL
 RODGAVCELOICGLAVADAGYSCVCGEERTSATLVKALPAKFTGLEGRNEAEVAGT
 AMLWELSKVAPVWRKGPENLRDGRYILRQEGTRCELOICGLAMADAGEYLVCVQ
 ERTSATLIRALPARFIEDVKNQAREGATVLOCELSAAPVWRKGRSRLDGRY
 SLRQGTKELOIRGLAVADAGEYSCVCGEERTSAMLTVALPFIKTEGLRNEATEG
 ATAVRCELSKQAPVWVKHETLRDGRHSLRQDARGCELOIRGLVAEADAGEYLWIC
 KERTSAMLTVRAMPKFIIEGRNEATEGDTATLWELSKAAPVWRKGRHETLRDGR
 RHLRQDGRCELOIRGLAVADAGEYSCVCGEERTSATLVRALPARFIEDVKNQEAR
 EGATAVLOCELSKAAPVWRKSETLRDGRYLRQDGRCELOICHLGVADAGEYSC
 VCGOERTSATLVRAPQVPREPQLOQAEBSGTATLQCELSPTATVWVSKGLOLQ
 ANGRERLOGCTAEVLQDLOREDTGEYTCGSOATSLTVAAPVRLRLOHQ
 BVDSGTHALCELSRAGASVWRKGSLOLPPCKAYQWQDGAALVLRGVQEDBAG
 DYTCDTGHQSMASLVVRPRPKTRKTLQSEGTGDIARLCCOLSDAESGAVWMLK
 EGVELHAGPYEMRSQATRELILHLEAKDTGEYACVTGQKTAASLRVTEPEVTV
 RGLVAEYTDATEDVEFCEVSRAGATQWMLQPLQSNSETEVAVRDGRHILRLK
 GVTEPDAGTVFHLGNHASRAQLTVRAPEVITILEPLQDVLQSEGQASFCQSLRSAG
 QEARWALGVLOANEMNDITVEOGLTHLLTHKVTLEDAGTVSFHGTCSRAQLV
 TAKNTVNLGNLENVALRGGALFECOLSQEVAHTWLLDDEPVRTSENAEVFFENG
 LRHLLLNKLPQDSCTVTLAGDNVTSALFTVRGHWLEILEPLKXNAVAGAQART
 CTSEAVPVGASVYINGAAVQDDSDVTADGSHQALLRSLRAGHAGEVTFACD
 AVASARLVGLDPPDEDAVHAHSHTVLSWAAPMDSGGGLGVRYRKBGATQ
 WRLCHELVGPECVVDGLPAGETVFRVAAVPGVAGEPVHLPQTVLRAEPVFPVQ
 PSAPQVQAAGVSDLELEVAEAGEVWIKEMRIQPGREFVWQSGQMLVWKG
 FTADQGEVHGLAAGSICPAATFOVALSPADVAPOPSLPPEAAQEGDLHLLEA
 LARVRSREPTLSISELPEEDGSRQLPOEAEEVAPDLSEGYSTADELARTGDADL
 SHTSDESRAGTSLVLYLKKAGRPCTSLAKVGAAPAPSPKPPQOQFPLAARVP
 LGDLSLDDGSPMDKAAVKKAAQKGVKVKMKQEGPFMSFTGDTBAQGDAR
 LECVVAADVARWLDKVELDTRHHHIDQDGLTCSLLIAGLRADAGCTCVQS
 NKQVCTHSAQCVVSGSESAESSGGLDADAFRAARLHRLFTKSPAEVDEBELF
 LSADGEPMECISKETPAPVVPPEPLSLTSDAAPVELTELQNEOVQDGPVPSFDCV
 VEOGRPVMECISKETPAPVVPPEPLSLTSDAAPVELTELQNEOVQDGPVPSFDCV
 VTGQMPSPVWFKDKLLEDDHYMINEDQGGHLLITAVPADMEVYRCLAENSNG
 VSTKABELVDLTSITDTRADATESSSYSAQGLSSRQEGTSTDSGQLPQVVE
 ELRDLQVAPTRLAKFQKVKVGPAPRLYWKQDQPLTASAHRTMDTKLILHLELIS
 VTRDSQYAAVTSNANGAAYSSARLVGRDEPEEPKPSADVHEQLVPPMLRFTPK
 KVKGSSTITSVKVGEVPVPTVHREEARGVLMIGPTPGTVASAAQSHLSLJ
 VROHQGTCTIASNAGAOGLCSALHVGSPKVEOEKQKEALISTFLOGTQALSA
 QGFOTSAFDLGGKKEEPLAAEALGHLSAEVGTETFLQKLSQITEMVSAKITQA
 KLOVPGGSDSDSKTPSAPRHCPSRPSSTIQSSSESDGDARGEIIDLIVVTADLV
 PLGAQAQALIREQYVEVLDAAHPLRWLVRTPKSPSPQGVWSFAYLDRKLSP
 EWGAAPPEPGEAEVSEYVQELLSQELSEQAFVEELQSHHQLHCHERCPH
 VPIAVAQKAVIFERNEDIGRHSFLQELQCCDTRDDVAMCFIKNOAAFEQGLFLV
 GRVQAEVYVSTAIQETFYKKAEEALLAGDPSPPPPLQHYLEQVRYQVQALLK
 ELIENKARNQNCALLQAVVAVSALPQRAENKLHVSLEMYPTQLALGEPTRQGHF
 IWMGAPQMPKGNHNVLPFRNLVICKPRDSRTDVTYVFRNMKLSIDLND
 QVEDDRAFEVQEREDSVRYLLQARTLIIKSSWKEICGIIQORLALPYRPPDRE
 ELACTABELGTYKLACRVGTGPKPVLSWYKDKGKAVQVDPHLIEDPDGSCALIDS
 LTVGDSQGLMCFASAAGNCTGLKILVQVFRPVNKLASPFVEGEDAQFTCTIEGA
 PYQIYRQKGLLITNKFTQSEPSGGLLVIRAAKEDJGLVECELVNLRGSRAR
 ASALRLIQSMLQAEQCHREQLVAEVEVTEQTKVTKTIVIEITTVTKSPQROG
 RSPKSPSRSPSRCSAPLPGGLIADLLYLPAGQPRRPEAPGPKVPVPTLYVTEA
 EAHSPALPGSPQPKWVEVEETEVAKMGKQGVSTTEVPSRSGHGLFTLPGATP
 GGDPSNNSNNKLQAQEAQGTAMGVREPLRVDRAGSVDMASGMSLEEGMT
 BEAGEEFGDGAFTVEESODTHLGDGRDPLKTHNGMLTLADLSDYVPEGETPHC
 GGPQGPAPDDPPCEVSVIQRIGEPTVQGVLLSVGHALGPRGLGFRPEPGASPP
 GPQVRSLEGTSFLREAPAPVGSAPWTQSFCTIRRSADSGQSSFTTELSTQTVNFG

| | | | |
|----------------------------|-----|---|------|
| ORIGIN | | TVGETVTLHLICPDRDGEAAQP" | |
| Query Match | | 100.0%; Score 1070; DB 9; Length 20435; | |
| Best Local Similarity | | 100.0%; Pred. No. 1.3e-119; | |
| Matches 1070; Conservative | | 0; Mismatches 0; Indels 0; Gaps 0; | |
| QY | 1 | TGCTTACAGAGCCACATCCGCGCGCTCCAGAGCCGCCATAGAGAGTCCCGG | 60 |
| DB | 1 | TGCTTACAGAGCCACATCCGCGCGCTCCAGAGCCGCCATAGAGAGTCCCGG | 60 |
| QY | 61 | CGCCACCGTATGATCAGCCACAGTTACGCGGGCGCCCGCTTCTTCCACCGGCCCA | 120 |
| DB | 61 | CGCCACCGTATGATCAGCCACAGTTACGCGGGCGCCCGCTTCTTCCACCGGCCCA | 120 |
| QY | 121 | AGGCTTCGTGGTGGGGAAGGAGCCACCTCAGCTGCGCAGATCGTGGGTATC | 180 |
| DB | 121 | AGGCTTCGTGGTGGGGAAGGAGCCACCTCAGCTGCGCAGATCGTGGGTATC | 180 |
| QY | 181 | CCAGCCACAGTGGGAGAGGACAGCAGCGGTGACCGCGCGCGCGCTTC | 240 |
| DB | 181 | CCAGCCACAGTGGGAGAGGACAGCAGCGGTGACCGCGCGCGCGCTTC | 240 |
| QY | 241 | CTCTGGCCAGGAGCGGACCTTACCGCTCCTATCTCTGGACCTCGCGTGGCGACA | 300 |
| DB | 241 | GTCTGGCCAGGAGCGGACCTTACCGCTCCTATCTCTGGACCTCGCGTGGCGACA | 300 |
| QY | 301 | GTGGCAATAGTGGCGCGCGCAATGCAATAGGAGGCTTTCGCTGCGTGGGCC | 360 |
| DB | 301 | GTGGCAATAGTGGCGCGCGCAATGCAATAGGAGGCTTTCGCTGCGTGGGCC | 360 |
| QY | 361 | TGCAGTGGAGCGGAGCGCGTGGCGGAGGAGCGCGCTTCTCTGCTGGGCCCA | 420 |
| DB | 361 | TGCAGTGGAGCGGAGCGCGTGGCGGAGGAGCGCGCTTCTCTGCTGGGCCCA | 420 |
| QY | 421 | CGTCATCCGCGTGGAGGGCTCAGAGGCCACTTCGCTGCGCGTGGTGGCTCC | 480 |
| DB | 421 | CGTCATCCGCGTGGAGGGCTCAGAGGCCACTTCGCTGCGCGTGGTGGCTCC | 480 |
| QY | 481 | CGAGCGCGGAGTGGCTGCTCAGGAGCGGCGCGCTGGGTGACCGCGCGCC | 540 |
| DB | 481 | CGAGCGCGGAGTGGCTGCTCAGGAGCGGCGCGCTGGGTGACCGCGCGCC | 540 |
| QY | 541 | CGCTGCGGTGGAGAGTCCGCGAGGCAAGTGGCTGCGCATTCGGCGCGCGCGCG | 600 |
| DB | 541 | CGCTGCGGTGGAGAGTCCGCGAGGCAAGTGGCTGCGCATTCGGCGCGCGCGCG | 600 |
| QY | 601 | CGAGCGCGGACTTACAGAGTCCGCGGAGAGCCGCTGGCGCTGCCAGCGCGCG | 660 |
| DB | 601 | CGAGCGCGGACTTACAGAGTCCGCGGAGAGCCGCTGGCGCTGCCAGCGCGCG | 660 |
| QY | 661 | CGGCGCTAGTGGTGGACTCGGACCGCGGACAGCGCGCGCGCGGACCTCCACGG | 720 |
| DB | 661 | CGGCGCTAGTGGTGGACTCGGACCGCGGACAGCGCGCGCGGACCTCCACGG | 720 |
| QY | 721 | CGGCGCTCTGGCGACCTGAGCGGCGCGGAGGCTATGCGCGCGGCGCGCGCG | 780 |
| DB | 721 | CGGCGCTCTGGCGACCTGAGCGGCGCGGAGGCTATGCGCGCGGCGCGCGCG | 780 |
| QY | 781 | CCTACCGCCAGACACCGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 840 |
| DB | 781 | CCTACCGCCAGACACCGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 840 |
| QY | 841 | TCAGTGTCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 900 |
| DB | 841 | TCAGTGTCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 900 |
| QY | 901 | TGGTACCGAGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 960 |
| DB | 901 | TGGTACCGAGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 960 |
| QY | 961 | AGATCCTCTTCTGCAAGCAGTTCGAGCCGCGCGCTCTACACCTGCGCGCTCC | 1020 |
| DB | 961 | AGATCCTCTTCTGCAAGCAGTTCGAGCCGCGCGCTCTACACCTGCGCGCTCC | 1020 |

| | | | |
|----|------|--|------|
| Db | 961 | AGATCCTCTTCTGCAAGCAGTTCGAGCCGCGCGCTCTACACCTGCGCGCTCCAACTCG | 1020 |
| QY | 1021 | TGGCCAGACCTACAGCTCTGTGCTGGTCTAGTGGCGGAGCCCGGTT | 1070 |
| DB | 1021 | TGGCCAGACCTACAGCTCTGTGCTGGTCTAGTGGCGGAGCCCGGTT | 1070 |

| | | | |
|------------|---|-------------|--------|
| RESULT 2 | | | |
| AL359510 | 95745 bp | DNA | linear |
| LOCUS | Human DNA sequence from clone Rp11-520H14 on chromosome 1, complete | | |
| DEFINITION | sequence. | | |
| ACCESSION | AL359510 | | |
| VERSION | AL359510.24 | GI:18496240 | |
| KEYWORDS | HTG. | | |
| SOURCE | Homo sapiens (human) | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | 1 (bases 1 to 95745) | | |
| AUTHORS | Blakey,S. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (02-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk | | |
| COMMENT | On Feb 4, 2002 this sequence version replaced gi:18121469. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chrl | | |
| | Rp11-520H14 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm | | |
| | VECTOR: pBACe3.6 | | |
| | IMPORTANT: This sequence is not the entire insert of clone Rp11-520H14. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone RP5-1139B12 is at 93746 in this sequence. The true right end of clone RP5-881P19 is at 2000 in this sequence. | | |

| | |
|--------------|---|
| FEATURES | Location/Qualifiers |
| source | 1. .95745 |
| | /organism="Homo sapiens" |
| | /mol_type="genomic DNA" |
| | /db_xref="taxon:9606" |
| | /chromosome="1" |
| | /clone="Rp11-520H14" |
| | /clone_lib="RPCI-11.2" |
| | /2018..20426 |
| misc_feature | /note="Sequence from uni-directional dGTP big dye terminator reads only." |
| misc_feature | 93348..93408 |
| | /note="Single clone region. Assembly confirmed by restriction digest data." |

| | |
|--------|--|
| ORIGIN | |
|--------|--|

93783 ACCTGACGCGGTCAAACCTCGTGGGCAGACCTACAGCTCTGTGCTGGTGTAGTGC 93842

```

119561 CGCGCTTCCGCGTGGCCGACGAGCGGAGTGTATACCGCTCACAATTTCTCGATCTGGCGC 119502
292 TGGCGGACAGTGGCAATAGTGTGCGCGCGCAATGCAATGAGGAGCGCTTGGCTG 351
119501 TAGTGTACAGCGGCGAGTGTGTGTGCGCGGAGGAGCAATAGGAGCGCTTGGCG 119442
352 CCGTGGGCGCTGAGGTGACGCGAGCGCGCGTGTGCGCGGAGGAGCGCGCTTGGCTG 411
119441 CCGTGGGCGCTGAGGTGAGCTGCGAGGCGCGTGTGCGCGGAGGAGCGCGCTTGGCTG 119382
412 TGGCGGCGCGCTGAGGTGAGGAGCTGCGGAGGCGCTGAGGAGCGCGCTTGGCTGCGGCTG 471
119381 TGGCGGCGCGCATTCATTCGCGTGTGCGGAGGCGCGAGCGCGCTTGGCTGCGGCTG 119322
472 GTGCGCTCCCGAGGCGCGAGTGTGCTGCTCAAGAGCGCGCGCGCTTGGCTGCGGCTG 531
119321 GAGGCTCACCAACCTGCTGTGAGCTGTGCTCAAGAGTGTGCGGCGCGCTTGGCTGCGGCTG 119262
532 ACGGCGCGCGCTGCGGTGTGAGGAGCTGCGGAGGCGAGTGTGCGCTGCGCTGCGGCTGCGGCTG 591
119261 ATGCGCGCGCGCTGCGGTGTGAGGAGCAACCGGCGGAGTGTGCGGCTTGGCTGCGGCTG 119202
592 CCGCGCGCGCGCGCGGCGCTTACAGAGTGTGCGCGCGGAGCAACCGGCGGCTTGGCTGCGGCTGCGCA 651
119201 CAAGGCTCGCGATGTGTGCGCGCTTACAGAGTGTGCGCGGAGCAATCGGCTGCGGCTTGGCA 119142
652 CGCGCGCGCGCGCGCTGTGTGAGTGTGCGAGCGCGCGGAGCAACCGGCGGCTTGGCTGCGGCTGCGG 711
119141 CGCGTGTGCGCGCGCTGTGTGAGTGTGCGAGCGCGCGGAGCAACCGGCGGCTTGGCTGCGGCTGCGG 119085
712 CCGCGCGCGCGCGCTGTGTGAGTGTGCGAGCGCGCGGAGCAACCGGCGGCTTGGCTGCGGCTGCGG 771
119084 CCGCGCGCGCGCGCTGTGTGAGTGTGCGAGCGCGCGGAGCAACCGGCGGCTTGGCTGCGGCTGCGG 119025
772 CGCGCGCGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 831
119024 GCATACCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 118965
832 ACGGCGCGCGCGCTGTGTGAGTGTGCGAGCGCGCGGAGCAACCGGCGGCTTGGCTGCGGAGG 891
118964 ACGGCGCGCGCGCTGTGTGAGTGTGCGAGCGCGCGGAGCAACCGGCGGCTTGGCTGCGGAGG 118905
892 ACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951
118904 ACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 118845
952 TCGTGTCTCAAGTCTTCTGCAAGCGAGTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1011
118844 TCGTGTCTTAAATCTTCTTCTGCAAGCGAGTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 118785
1012 CCAACCTGTGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1059
118784 CCAACCTGTGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 118737

```

RESULT 4

```

AC099089/c
LOCUS AC099089 260998 bp DNA linear HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-154E3, WORKING DRAFT SEQUENCE.
ACCESSION AC099089
VERSION AC099089.5 GI:30522701
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Rattus.
REFERENCE 1 (bases 1 to 260998)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

```

Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,C., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Dragan,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gregoridis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hernandez,J.,
 Harvey,P., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.B., Hodgson,A., Hogues,M.,
 Hollins,B., Howells,S., Huiyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowals,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorusnewa,L., Louised,H., Lorado,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mathew,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
 Nwaokemelen,O., Okwodu,G., Olarunpunsagoon,A., Pal,S., Parks,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
 Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soosa,J.,
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstein,G. and Gibbs,R.A.

TITLE

Direct Submission

Unpublished

2 (bases 1 to 260998)

Direct Submission

Submitted (09-NOV-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 260998)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center

| | | | | | |
|------------|--|---|------|---|------|
| D | b | | 965 | CGTGTCAAGGTTCTACTCCAGGCCAAGATCGTGGCTCTACGTCGCGCGCGC | 1024 |
| Q | y | | 1013 | CAACCTCGTGCCGACAGCTACAGCTGTGTGTCGTGTCGTAGTCGCGAGCCC | 1064 |
| D | b | | 1025 | CAACTCGCGGGCCAGACGCTCAGTGCCTGTCAGCTGCACGTGAAGAAGCCC | 1076 |
| | | | | | |
| R | E | S | U | L | T |
| RESULT | 11 | | | | |
| LOCUS | AC009955/c | | | | |
| DEFINITION | Homo sapiens BAC clone RP11-256f23 from 2, complete sequence. | | | | |
| ACCESSION | AC009955 | | | | |
| VERSION | AC009955.4 GI:9581957 | | | | |
| KEYWORDS | HTG. | | | | |
| SOURCE | Homo sapiens (human) | | | | |
| ORGANISM | Homo sapiens | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 135033) | | | | |
| AUTHORS | Sulston,J.E. and Waterston,R. | | | | |
| TITLE | Toward a complete human genome sequence | | | | |
| JOURNAL | Genome Res. 8 (11), 1097-1108 (1998) | | | | |
| MEDLINE | 99063792 | | | | |
| PUBMED | 9847074 | | | | |
| REFERENCE | 2 (bases 1 to 135033) | | | | |
| AUTHORS | Du,H., Maupin,R., Hawkins,M. and Hodges,J. | | | | |
| TITLE | The sequence of Homo sapiens BAC clone RP11-256f23 | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 3 (bases 1 to 135033) | | | | |
| AUTHORS | Waterston,R.H. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (08-SEP-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA | | | | |
| REFERENCE | 4 (bases 1 to 135033) | | | | |
| AUTHORS | Waterston,R.H. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (29-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA | | | | |
| REFERENCE | 5 (bases 1 to 135033) | | | | |
| AUTHORS | Waterston,R.H. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA | | | | |
| REFERENCE | 6 (bases 1 to 135033) | | | | |
| AUTHORS | Waterston,R. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (14-OCT-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA | | | | |
| REFERENCE | 7 (bases 1 to 135033) | | | | |
| AUTHORS | Waterston,R. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA | | | | |
| COMMENT | On Jul 29, 2000 this sequence version replaced gi:8569874. | | | | |

| | | | |
|----|-------|--|-------|
| Db | 42125 | CGCATGAAGCGAGCTCGGGATCAGGGAGAGCCCGCTTCTCGCTCCGCG | 42066 |
| Qy | 122 | GGCTTCGTGTCGTGGGCAAGGACCCCTCAGCTGCAGATCGTGGTATCC | 181 |
| Db | 42065 | GCCTGTGGGTGTAAAGTGGCCGAGGCGGAGCTCAAGTGTGTCTGGGAGCC | 42006 |
| Qy | 182 | CACGCCACAGTGTGGAGAGGACACGAGCGGTGACGCGCGCGCGCTTCGG | 241 |
| Db | 42005 | GCCGCTGTAGTGTGGAGAGGCGGAGCAGCTGGCGCTCGGAACGCTGAG | 41946 |
| Qy | 242 | TCTGGCCAGGACGCGACCTTACCGCTCACTATCTCGACTGCGCTGGGACAG | 301 |
| Db | 41945 | CTTCCCGCGAGCGCGGACGCGCTTCTGCTGACGCGCGCACTGCCACCGACG | 41896 |
| Qy | 302 | TGGCAATACGTGTGCGCGCGCAATGCATAGGCGAGCGCTTCTCGCTGGCGCT | 361 |
| Db | 41895 | GGGGTCTACGTGTGCGCGCGCGACGCGCGGAGCGCTACGCGGGCGCGCT | 41826 |
| Qy | 362 | GCAGGTGAGCGCGAGCGCGGTGCGCGGAGCGCGGACCTTCTGTGCGGCCAC | 421 |
| Db | 41825 | CACCGTGTGAGCGCGCGCTCCGACCGCGAGCTCAGCGCGCGCGCTGCC | 41766 |
| Qy | 422 | GTCCATCGGTGCGCGAGGCTCAGAGGCGACCTTCGCTGCG-----CGTGGTGG | 475 |
| Db | 41765 | ATCCCGGGTCCGGAGGCGCGCGCTTCTCTCAGGGGCTCGATCCAGTGGT | 41706 |
| Qy | 476 | CTCCCGAGCGCGAGCTGCTGCTCAAGAGCGCGCGCTGGTGTGAGCCCGACGG | 535 |
| Db | 41705 | GCTCGGGGCGGAGTGTGTGCTGCTGCGCGCGCGCGCTCCCGAGCCACACT | 41646 |
| Qy | 536 | CCCCCGGTGCGGTGAGGAGCTCGGAGGCGAGTGTGCTGCGCATTCGGGCGCGG | 595 |
| Db | 41645 | GTACTGGAGAGGACCGGATGCGCTGGAGAGTGTGGACAGCAGCCTTCGCGCT | 41586 |
| Qy | 596 | GCCCGCGAGCGCGGCTTACGAGGTCGCGCGCGAGAACCGCTGGGC--GTTGCCAGC | 653 |
| Db | 41585 | CCAGCGCGCGCGCGGAGCGCGCGCGCGAGCTGCGCATCTGGCGGCG | 41526 |
| Qy | 654 | GCGCGCG--CGCGGTAGTGTGAGTGTGAGCGCGCGAGCACGCGCGCGCGGAC | 712 |
| Db | 41525 | TCGCTGCGGATTCGCGGCTTACGTGTGCGCGCGCGCGCGCGCGCGCGCA | 41466 |
| Qy | 713 | CTCCAGCGCGGCTCTCGCGCGCACCTGCGCGCGCGCGAGGCTATGCGCGCGAGGG | 772 |
| Db | 41465 | GCGGGGCGGTGTTCAGGTGTCACGAGCGCGCGCGCGCGCGCGCGCGCGACGA | 41406 |
| Qy | 773 | CGCCCCCGCTC-----ACCGCCGAGCACCGCGCGCGCGCGCGCGCGCGCTGA | 823 |
| Db | 41405 | GCGCGCGCGCGGTGGAGCGGCTCAAGTGGCGGCTTAAGACCTTCTGGGTGAACGA | 41346 |
| Qy | 824 | AGGCAAGCAGCGCGCTCAGCTGCTAGTACCGCGAGCGCGCGCGCGCGCGCGGTG | 883 |
| Db | 41345 | GGCAAGCAGCGATG | 41286 |
| Qy | 884 | GAAGAAGCAGCA | 943 |
| Db | 41285 | GCACGTGGAGGCGCGCGCGCTGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 41226 |
| Qy | 944 | GGAGACTGTGCTCAAGATCTCTTCTGCAAGAGTGTGACCGCGCGCGCTTACACTG | 1003 |
| Db | 41225 | CGCGCGCTTCTGCTCAAGTGTCTTACTGCGCGCGCGCGCGCGCGCGCGCGCTG | 41166 |
| Qy | 1004 | CAGCGCGTCAAGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG | 1053 |
| Db | 41165 | GCG | 1063 |
| Qy | 1064 | GCGCG 1068 | |
| Db | 41105 | GCGCG 41101 | |

RESULT 12

AC040991
LOCUS
DEFINITION

186883 bp DNA linear HTG 24-AUG-2002
Homo sapiens chromosome 2 clone RP11-123E11 map 2, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
AC040991
HTG; HTGS_PHASE1; HTGS_DRAFT.
AC040991.2 GI:8014564
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 186883)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 2, clone RP11-123E11
Unpublished
2 (bases 1 to 186883)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Corry,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 186883)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Corry,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 22, 2000 this sequence version replaced gi:7534193.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L9492
 Center clone name: 123_E.11

----- Summary Statistics

Sequencing Vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 175185 bases at least Q40
 Consensus quality: 181490 bases at least Q30
 Consensus quality: 183848 bases at least Q20
 Insert size: 190000; agarose-fp
 Insert size: 185283; sum-of-contigs
 Quality coverage: 4.3 in Q20 bases; agarose-fp
 Quality coverage: 4.4 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1462: contig of 1462 bp in length
 * 1463 1562: gap of 100 bp
 * 1563 4486: contig of 2924 bp in length
 * 4487 4586: gap of 100 bp
 * 4587 9251: contig of 4665 bp in length
 * 9252 9351: gap of 100 bp
 * 9352 12768: contig of 3417 bp in length
 * 12769 12868: gap of 100 bp
 * 12869 17380: contig of 4512 bp in length
 * 17381 17480: gap of 100 bp
 * 17481 22333: contig of 4753 bp in length
 * 22334 22334: gap of 100 bp
 * 22334 25600: contig of 3267 bp in length
 * 25601 25700: gap of 100 bp
 * 25701 31826: contig of 6126 bp in length
 * 31827 31928: gap of 100 bp
 * 31927 41356: contig of 9430 bp in length
 * 41357 41456: gap of 100 bp
 * 41457 53195: contig of 11739 bp in length
 * 53196 53295: gap of 100 bp
 * 53296 64559: contig of 11264 bp in length
 * 64560 64559: gap of 100 bp
 * 64560 79824: contig of 15165 bp in length
 * 79825 79924: gap of 100 bp
 * 79925 99903: contig of 19985 bp in length
 * 99910 100009: gap of 100 bp
 * 100010 121977: contig of 21968 bp in length
 * 121978 122077: gap of 100 bp
 * 122078 143182: contig of 21105 bp in length
 * 143183 143284: gap of 100 bp
 * 143285 164587: contig of 21305 bp in length
 * 164588 164587: gap of 100 bp
 * 164588 186883: contig of 22196 bp in length.
 Location/Qualifiers
 1. 186883
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-123E11"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 1462
 /note="assembly_fragment"
 1583. 4486
 /note="assembly_fragment"
 4587. 9251
 /note="assembly_fragment"
 9352. 12768

FEATURES
 source

misc_feature
 /note="assembly_fragment"
 12869. 17380
 /note="assembly_fragment"
 clone_end:17
 vector_side:left
 17481. 22333
 /note="assembly_fragment"
 22334. 25600
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right
 25701. 31826
 /note="assembly_fragment"
 31927. 41356
 /note="assembly_fragment"
 41457. 53195
 /note="assembly_fragment"
 53296. 64559
 /note="assembly_fragment"
 64660. 79824
 /note="assembly_fragment"
 79925. 99909
 /note="assembly_fragment"
 100010. 121977
 /note="assembly_fragment"
 122078. 143182
 /note="assembly_fragment"
 143283. 164587
 /note="assembly_fragment"
 164588. 186883
 /note="assembly_fragment"
 ORIGIN
 Query Match .. 18.5%; Score 197.8; DB 2; Length 186883;
 Best Local Similarity 51.6%; Pred. No. 1.4e-15;
 Matches 560; Conservative 0; Mismatches 507; Indels 18; Gaps 4;
 QY 2 GCCTACAGGAGCCACACTCCGGCCGTCGCCAGAGCCGCCATAGAGAGGTCCTCCG 61
 Db 109463 GCTGCCCTCCGACACCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 109522
 QY 62 CGCCACCTCATGTGATCAGCCACAGTTTCAGCGGGGCGCCCGCTTTCTACCCCGGCCAA 121
 Db 109523 CGCGATGAAGGCGAGCTCGGGGATCAGGGAGCCCGCTGCTTCTCGCTTCCGCG 109582
 QY 122 GGCCTTCGTGTGTGCGTGGGCAAGGAGCCACCTCAGCTGCCAGATCGTGGGTATCC 181
 Db 109583 GCCTGTGCGGTGTGTAAGTGGCGCGGAGCGGAGCTCAAGTGCCTGCTTCTGGGGAGCC 109642
 QY 182 CACGCCACAGTGAGCTGGGAGAGGACACAGCAGCCGCTGACGCCCGCGCGCTTCCTCG 241
 Db 109643 CGCGCTGTACTGTGTGGGAGAGGGGGGCGGAGCTGGCGGCTCGGAGCGCTGAG 109702
 QY 242 TCTGGCCAGGAGCGGACCTCTTACCGGCTCACTATCTGTGACCTGGCGCTGGCGGACAG 301
 Db 109703 CTTCCCGCGGAGCGGCGGAGCAGCGGCTGCTGTACCGCGGAGCTGCCACACGCG 109762
 QY 302 TGGGCAATACGTGTGCGCGCGCAATGCCATAGGAGAGGCTTCGCTGCCGTGGGCGCT 361
 Db 109763 GGGGGTCTACGTGTGCGCGCGCCCAAGCGCGCGGCGGAGGCTTACGCGGCGCGCGCT 109822
 QY 362 GCAGGTGAGCGCGAGCGCGGTGCGCGGAGCGCGGAGCGCGGAGCTTCTGTGGGCGCCAC 421
 Db 109823 CACCGTGTGTGAGCGCGCGGCTCCGACCCCGAGCTCAGCGCGCGGAGCGCGGCTGCC 109882
 QY 422 GTCCATCGCGTGTGCGGAGGCGCTCAGAGGCGCCCTTCCTCGCTGCCG 475
 Db 109883 ATCGCGGGGTCCCGGAGGCGCGCGCGCTTCTTCTCAGCGGCGCTCGATCCAGTGGT 109942
 QY 476 CTCGCCGAGGCGCGAGTGTGTTCAAAGGAGCGCGCGCGCTGGGTGAGCCCGACGG 535
 Db 109943 GCTCGGGGGCGGAGGTGTGTGCTGACGTGCGGGCGGGGGGCTCCCGAGGCCCACT 110002

misc_feature
 misc_feature
 misc_feature
 misc_feature

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., L. Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensu, H., L., Louised, H., Lozardo, R. J., Lu, X., Ma, J., Maheshwar, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathewney, S., McLeod, M. P., McNeill, T. Z., Meenen, B., Milosavljevic, A., Miner, G. S., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankevris, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaekalameh, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfanckoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Fu, L., L., Puzao, M., Qiuoz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Rodley, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, D., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vexa, V., Villasanua, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willison, R., Wlacyk, R., Woden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 225471)
Worley, K.C.
Direct Submission
Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 225471)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

Baylon, Pizarra, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:22855903.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:22856903.

and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, scaffolds are numbered and ordered as they were connected.

assembly 1a 'contig-scaffolds' 7: within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```

contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- "Project Information
Center project name: GSN
Center clone name: CH230-61F12
----- Summary Statistics
Assembly program: Phrap; version 0.90329
Consensus quality: 194781 bases at least Q40
Consensus quality: 197982 bases at least Q30
Consensus quality: 200161 bases at least Q20
Estimated insert size: 194216; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

```

Assembly program: Parap; version 0.90329
 Consensus quality: 194781 bases at least Q40
 Consensus quality: 197982 bases at least Q30
 Consensus quality: 200161 bases at least Q20
 Estimated insert size: 194216; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation


```

*****
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 224074: contig of 224074 bp in length
* 224075 224174: gap of unknown length
* 224175 225471: contig of 1297 bp in length.
FEATURES             source
     Location/Qualifiers
         1..225471
             /organism="Rattus norvegicus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10116"
             /clone="CH230-61F12"
         1..1154
             /note="wgs contig"
         17776..18866
             /note="wgs contig"
         134828..136148
             /note="wgs contig"
         222128..224074
             /note="wgs contig"
misc_feature
misc_feature
misc_feature
misc_feature
ORIGIN
Query Match          16.9%; Score 181; DB 2; Length 225471;
Best Local Similarity 52.18; Pred. No. 1.4e-13;
Matches 509; Conservative 0; Mismatches 450; Indels 18; Gaps 4;
QY 97 GCCTCCGCTTCTCACCAGGCGCCAGGCTTGTGTTGTCGTTGGGAGGAGGCGCACCC 156
Db 142585 CCCCCTGTTTCTACCTAGCTTCCCGCGGCGCTGCGGCTGTTAAAGTGGCGGAGCGCGAGC 142644
QY 157 TCAGCTGCAGATCGTGGGTAATCCACGCGCACAGGTGAGCTGGAGAGGAGCAGCAGC 216
Db 142645 TCAATGGTGGTCTCTGGAGAGCGCGCGCCACCGCTCTTGTGGAGAGAGGGCGGAGC 142704
QY 217 CGGTAGCGCGCGCGCGCTTCCGTCTGCGCCAGGAGCGGACCTTACCGCTCACTA 276
Db 142705 AGCTGGTGGCTCGAGAGCGCTGAGCTTCCAGAGGAGCGCGCGCTGCTGT 142764
QY 277 TCCTGGACCTTGGCGCTGGGCGAGAGTGGGCAATACGTGTGCGCGCGCGCAATGCCATAG 336
Db 142765 TGACGGGGGCTGCTCCACAGAGCGCGGGTCTACGTGTGCGCGCGCGCGCGCGCG 142824
QY 337 GCGAGGCGCTTCTC-----GCTCCGCTGGGCTGAGGTGAGCGCGGAGCGCGTGGCGCG 390
Db 142825 GAGAGGCGCTACGCGCGCGCGCTGCTGCTGGAACCCCGCGCGCGCGAGCAGCGGAGC 142884
QY 391 AGCAGGCGCGCTTCTCTGCGCGCGCGCTTCCATCCGCGTGGCGGAGGCTCAGAGG 450
Db 142885 CCCAGTCTTCGAGTGTCCCGCGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG 142944
QY 451 CCACCTTCCGCTGCGCGTGGGCTTCCCGAGGCGCGCGAGTGGTGTGCTCAAGGAGC 510
Db 142945 TGACGGAGCCCGAGTCCAGTCCAGTGGTGTGCGAGGCGGAGGTGGTGTGAGTGCAGG 143004
QY 511 GGCGCGCGCTGGTGTGAGCGCGAGCGCGCGCGCGCTGCGCGTGGAGGAGCTCGCGGAGGCA 570
Db 143005 TGGAGGCGCTTCCCGGAGCGCGCGCGCGCTGCTGCGGAGGAGGATGGGCTTGGAGCAG 143064
QY 571 GTGCGCTGCGCATTCGGGCGCGCGCGCGCGCGCGCGCGCTTACGAGGCTCCGCGCGCG 630
Db 143065 TGTGGGACAGAGCGCACTACAGCTGGAGCGGAGCGCGCGCGCGAGTGGAGCGGCGCA 143124
QY 631 AGAACCGCGTGGCG--GCTGCGAGCGCGCGCG--CGGCGGTAGTGGTGGAGCTCGGAGCGCG 687
Db 143125 GCCTGACGTTACGATCTTGGCAGCGCGCGCTGCGCGGATTCGCGGGGTGTACGTGTGCCAG 143184

```

```

QY 688 CGGACACGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
Db 143185 CCGGCAACGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143244
QY 748 GCGCGAGGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 798
Db 143245 AGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143304
QY 799 GCACGCGCAGCTGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 858
Db 143305 CGCCCAAGACCTTCTGGGTGAACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143364
QY 859 GCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 918
Db 143365 GCAAGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143424
QY 919 GCACGCTGCTGACGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 978
Db 143425 GCGCGCTCATGTACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143484
QY 979 AGTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1038
Db 143485 CCAAGGACCGTGGCGCTTACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143544
QY 1039 CTGTGCTGCTGCTAGTG 1055
Db 143545 CGTTTCAGCTGCACGTG 143561

```

RESULT 14

AC121633/c

LOCUS

DEFINITION

Rattus norvegicus clone CH230-62C23, WORKING DRAFT SEQUENCE, 6

unorderd pieces.

ACCESSION

AC121633

VERSION

AC121633.4 GI:25137860

KEYWORDS

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus

REFERENCE

1 (bases 1 to 266167)

Munz, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anvalbechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Chen, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,

Hernandez, J., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hui, X., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kovis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorenshew, L., Loulseg, H., Lozano, R. J., Lu, X., Ma, J.,

Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poinexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sleson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Staimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Submitted (21-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 266167)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:22856334.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('a' contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GWAS
Center clone name: CH230-62C23
Assembly program: Phrap; version 0.990329
Consensus quality: 252717 bases at least Q40
Consensus quality: 255419 bases at least Q30
Consensus quality: 257320 bases at least Q20
Estimated insert size: 258645; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will

* be preserved.
* 1 256774: contig of 256774 bp in length
* 256775 256874: gap of unknown length
* 256875 258093: contig of 1219 bp in length
* 258094 258193: gap of unknown length
* 258194 259332: contig of 1339 bp in length
* 259333 259832: gap of unknown length
* 259833 261419: contig of 1787 bp in length
* 259633 261519: gap of unknown length
* 261420 262537: contig of 1018 bp in length
* 262538 262637: gap of unknown length
* 262638 266167: contig of 3530 bp in length.
FEATURES
source
Location/Qualifiers
1..266167
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-62C23"

ORIGIN

Query Match 16.9%; Score 181; DB 2; Length 266167;
Best Local Similarity 52.1%; Pred. No. 1.3e-13; Indels 18; Gaps 4;
Matches 509; Conservative 0; Mismatches 450;
QY 97 CGCCCCGCTTTCTCACC CGGCCCAAGCCCTTCGTGTCGCTGGGCAAGGACGCCACCC 156
Db 244163 CCCCCGTTTCTTACGCTTCCCGCGCGCGTGTGCGGTAAAGTGGCGGAGGCCGAGC 244104
QY 157 TCAGCTGCAGATCGTGGGTAAATCCACGCCACAGTGAGCTGGGAGAGGACGACGAC 216
Db 244103 TCAAAATGCTGTGCTCTGGGAGAGCGCCGCCACCTCTTGTGGGAGAAAGCGGCGAC 244044
QY 217 CGGTGACGCGCGCGCGCTTCCGTCGTGGCCAGGACGCGCACCTCTACCCGCTCACTA 276
Db 244043 AGCTGTGCGCTTCGGAGCGCTGAGCTTTCAGAGAGCGCGCCGAGACGCGCTGCTGT 243984
QY 277 TCTTGACCTGCGCTGGCGCGACAGTGGGCAATACGTGTGCCCGCGCGCAATGCGCATAG 336
Db 243983 TGAGCGCGCGTGTGCCACCGACCGCGGGTCTACGTGTGCCCGCGCGCAACCGCGCGC 243924
QY 337 GCGAGCGCTTC-----GCTGCGCTGGGCTCAGGTGACACCGGAGGCGGTCGCGCG 390
Db 243923 GAGAGCGCTTACGCGCGCGCGCGTCACTGCTGTAACCCCGCGCGCCAGAGCGCGAGC 243864
QY 391 AGCAGCGCGCGCACTTCTTCTGCTGCGGCCCACTCCATCCGCTGCGCGAGGCTCAGAG 450
Db 243863 CCCAGTCTTCGAGTGTCCGCGCGCGCTCCGCGCACCGGGAGGCGCGCCCGGTGTTTC 243804
QY 451 CACCTTCGCTGCCGCTGGGTGGCTCCCGAGCGCGCAGTACGTGTCAGGAGCG 510
Db 243803 TGACGGGACCCAGTCCAGTGGGTGCTCAGAGGCGCGAGGTGTGCTGACGTGCCAGG 243744
QY 511 GCGCGCGCTGGGTGAGCCCGACGCGCGCGCTGCGCTGGAGGAGCTCCGCGAGGCGAA 570
Db 243743 TGGAGGCTTCGCGAGCCCAAGCTGTACTGGGAGAGAGATGGATGGCTTTGGACGAAG 243684
QY 571 GTGCGCTGCGCAATTCGGGCGCGCGCGCGCGCGCGCTTACAGAGTTCGCGCGCG 630
Db 243683 TGTGGGACAGACGCCACTACAGCTGGAGCCGCGCGCGCGCGAGTGCAGCGGCGCAA 243624
QY 631 AGAACCGCTGGGC--GCTGCCAGCGCGCGCG--CGGCGCTAGTGTGGGCTCGGAGCGCG 687
Db 243623 GCCTGACGTTTACGATCTTCGGAGCGCGGCTGCCCGATTCGGGGTGTACGTGTGCACG 243564
QY 688 CGGACACGCGCAGCGCGCGCGCGCGCGCTCCACCGCGCGCTCTTGGCGCACTTCAGCGGC 747
Db 243563 CCGCGAAACCGCACGCGCCACGCGAGCGCGCGCGCGCTGTCTCCAGTGTGCACGACGCCACG 243504
QY 748 GCGCGAGGCTATGCGCGCGGAGGCGCGCGCG-----CCTCACCGCGCGACCG 798
Db 243503 AGAACCGCGCGCGAGGATCCGATAGCCCCCGGTACGCGTATAGAGCGCGCTCAATGCG 243444
QY 799 GCACGCGCACTGCACGCTGACTGAAGGCAACGACGCGCGCGCTCAGCTGCTACGTGACCG 858

Db 243443 CGCCCAAGACCTCTGGGTGAACGAGGCGACGACGCAAGTTCGGTGTACGTGTGG 243384
 QY 859 GCGAGCCCAAGCCGAGCGGTGTGGAGAAAGAGCGCCAGCTGGTGAACGAGCGCGCC 918
 Db 243383 GCAAGCCCTGAGCCCGAGATCGAATGGCACTTGGAGGCGCGCCCTCTCTCCCGACGCGC 243324
 QY 919 GCCAGCTGGTGTACGAGGACCGGAGGAGAACTTGGTGTCAAGATCCTTCTTGCAGGC 978
 Db 243323 CGCGCTCATGTACCGGACCGTACGCGTGTGTACTAAGTGTCTTACTTGCAGG 243264
 QY 979 AGTCGACCGCGGCTCTACACCTGCGACGGCGTCCAACTGTGGCGACGACCTACAGCT 1038
 Db 243263 CCAAGGACCGTGGGCTCTACGTGTGCGGCGCGCAACTCGCGGCGCAGACCTTAAGCG 243204
 QY 1039 CTGTGCTGGTGTAGTG 1055
 Db 243203 CGGTACGTCGACGTG 243187

RESULT 15
 AC115011 187999 bp DNA linear HTG 17-DEC-2003
 LOCUS Mus musculus chromosome 1 clone RP24-192A11 map 1, *** SEQUENCING
 DEFINITION IN PROGRESS **, 7 unordered pieces.
 ACCESSION AC115011
 VERSION AC115011.8 GI:39979476
 KEYWORDS HTG; HTGS PHASE1; HTGS FULLTOP; HTGS_ACTIVEFIN.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus chromosome 1, clone RP24-192A11
 1 (bases 1 to 187999)
 2 (bases 1 to 187999)
 Unpublished
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Bouckhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand, P., Jones, C.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Lacombe, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L.,
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
 Seaman, S., Sever, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (14-MAR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE 3 (bases 1 to 187999)
 AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
 Anderson, M., Atchachi, H. M., Barna, N., Bastien, V., Bloom, T.,
 Bouckhalter, B., Bouckhalter, B., Camarata, J., Chang, J., Choepel, Y.,
 Collymore, A., Cook, A., Cooke, P., Coris, B., Dearellano, K.,
 Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S.,
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, B., Hagos, B.,
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C.,

Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
 Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Schauer, S., Schupback, R., Seaman, S., Sever, P., Smith, C.,
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
 Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL COMMENT

Direct Submission
 Submitted (17-DEC-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Dec 17, 2003 this sequence version replaced gi:28927805.
 All repeats were identified using RepeatMasker.
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L24606
 Center clone name: 192_A_11

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 66296: contig of 66296 bp in length
 * 66297 66396: gap of 100 bp
 * 66397 98400: contig of 30004 bp in length
 * 98401 98500: gap of 100 bp
 * 98501 102462: contig of 5962 bp in length
 * 102463 102562: gap of 100 bp
 * 102563 122984: contig of 20422 bp in length
 * 122985 123084: gap of 100 bp
 * 123085 123388: contig of 19304 bp in length
 * 123389 142489: gap of 100 bp
 * 142490 144632: contig of 2143 bp in length
 * 144633 147311: gap of 100 bp
 * 147312 187999: contig of 43268 bp in length.

FEATURES
 source
 1. 187999
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /map="1"
 /chromosome="1"
 /clone="RP24-192A11"
 /clone_lib="RPCI-24 Male Mouse BAC"

ORIGIN

Query Match 15.6%; Score 166.8; DB 2; Length 187999;
 Best Local Similarity 51.2%; Pred. No. 7.2e-12;
 Matches 501; Conservative 0; Mismatches 457; Indels 20; Gaps 4;
 QY 97 CGCCCGCTTCTTCTCACCAGCGCCCGCGCTTCTGTTGTGCGTGGGCAAGACGCCACCC 156
 Db 78061 CCCCGTGTTCCTACGCTTCCCGCGGCCCGTGGTGGTGAAGTGGAGCGGCGGAGC 78120
 QY 157 TCAGTGTCCAGATCGTGGGTATATCCACGCCACAGGTGAGTGGAGGAAGACGAGCAGC 216
 Db 78121 TCAATGCGTGTGCTCTCGGAGAGCGCGGCCCACTGTCGTGTGGGAGAAAGCGGGGAGC 78180
 QY 217 CGGTGACGCGCGCGCGCGCTTCCGTCTGGCCACGAGGACGGCAGCTTACCGCTCACTA 276
 Db 78181 AGCTGTGGCTCTCCGAGCGCTAGCTTTCGGGAGGACGGCGCGGAGCAGCGCTGTCTG 78240

| | | | |
|----|-------|--|-------|
| QY | 277 | TCCTGGACATGGCGCTGGGGCGACAGTGGGGAATACTGTGCCGCGCGCGCAATGCCATAG | 336 |
| Db | 78241 | TGAGCGCGCGCTGCCCAACGACCGCGGGGTCTACGTGTCCGCGCGCGCAACGCGGCGC | 78300 |
| QY | 337 | GCAGAGGCTTTC-----GCTGCCGTGGGGCTGCAGGTGACCGGGAGGCCCGTGCGCCG | 390 |
| Db | 78301 | GAGAGGCTTACGGCGGCGCGCGCTCACGTGTGGAACCCCGCGCCCCGAGACCGGAGC | 78360 |
| QY | 391 | AGCAGGCGCGCACTTCTGCTTGGCCCAACGCTCCAATCCGCTGTGGCGAGGGCTCAGAGG | 450 |
| Db | 78361 | CCGAGTCTCTCCGAGTGTCCGCTCCCAACACCGGGCACCGGGAGGGCGCCCGAGATTCC | 78420 |
| QY | 451 | CCACCTTCGCTGCCGCGTGGTGTCTCCGAGGCCGACGACGTGAGCTGTGTCCAGGACG | 510 |
| Db | 78421 | TGACGGGGCCCCAGTCCCACTGGGTGCTCGAGGGGAGAGTGTGTGACGTGCCAGG | 78480 |
| QY | 511 | EGGGGCGCTGGGTGAGCCCGACGGCCCCGCGTGCCTGAGAGAGCTCGCGAGGCA | 570 |
| Db | 78481 | TGGAGGCTTTCGGAGGCCCAAGCTACTTGGGAGAGGATGGGATGGCTTGNACGAG | 78540 |
| QY | 571 | GTGGCTGCGCATTTCCGGCGCGCGCGCGCGGCGAGCGGGCACTTACGAGTTCGCGCGC | 630 |
| Db | 78541 | TATGGGACAGCAGCCACTTCAAGCTGGAGCCCGCGCGCGAGTACAGAGGGCGGA | 78600 |
| QY | 631 | AGAACCGCTGGCGCTGCCAGCGCGCGCGCGCTA---GTGGTGGACTTCGACGCC | 686 |
| Db | 78601 | GCTGACGTTGCCG-ATCTTGGCGGACCGGCTGCCCGATTCCGGGGTGTACTGTGTAC | 78659 |
| QY | 687 | CGGACACGCCACCGCGCGCGGACCTCACGGCGCGCTCTTGCGCGACCTGCAGCGG | 746 |
| Db | 78660 | GCCCGCAACGCGACCGCACGGCGAGCGGGCGCGTCTCCAGGTGCACCGCCCGC | 78719 |
| QY | 747 | CGCGGGAGGCTATGCGCGCGGAGGGCGGCCCGCGCTC-----ACCGCCAGCAC | 797 |
| Db | 78720 | GAGAGCCCGCCAGGATCCCGATGAGAACCCCAACCCCGTGTGGAGCCGCTCAAAGC | 78779 |
| QY | 798 | GGCACGGCACCTTGCAGGTGACTGAAGGCAAGACGCGCGCTCAGCTGTACTGTACC | 857 |
| Db | 78780 | CGCGCCNAGACTTCTGGGTGAACGAGGGCAAGACGCCAATTCGCTGTACTGTATG | 78839 |
| QY | 858 | GGCGAGCCCAAGCCCGAGACGGTGTGGAAGAGGACGGCCAGCTGGTACCGAGGCGG | 917 |
| Db | 78840 | GGCAACCTTGAGCCCGAGATCGAATGGCACTTGGAGGGCGGCCCTCTCTCCCCGATCG | 78899 |
| QY | 918 | CGCCACGTGGTGTACGAGACCGCGAGGAGAACTTCGTGCTCAAGATCCTCTTCTGCAAG | 977 |
| Db | 78900 | CGCGGCTCATGTACCGCACCGCGCGCTTGTACTTAAGGTGCTCTACTGCCAG | 78959 |
| QY | 978 | CAGTCTGGACCGCGGCTCTACACTCTCAGCGGTCCAACCTCGTGGGGCAGACTTACAGC | 1037 |
| Db | 78960 | GCCAAGGACGTTGGGCTCTACGTGTGCGGGCGCGAACTCGCGGGCCAGCCCTAAGC | 79019 |
| QY | 1038 | TCGTGCTGGTCTGATG | 1055 |
| Db | 79020 | CGGTCACGTGCACGTG | 79037 |

Search completed: September 19, 2004, 19:26:59
Job time : 4372.03 secs

This Page Blank (uspto)